AD_____
(Leave blank)

Award Number:
W81XWH-08-1-0720

TITLE:
Engineering Improvements in a Bacterial Therapeutic Delivery System for Breast Cancer

PRINCIPAL INVESTIGATOR:
Michael McClelland Ph.D.

CONTRACTING ORGANIZATION:
Vaccine Research Institute of San Diego
San Diego, CA 92121-1130

REPORT DATE:
September 2010

TYPE OF REPORT:
Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: (Check one)

X Approved for public release; distribution unlimited

☐ Distribution limited to U.S. Government agencies only; report contains proprietary information

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
01/09/2010	Annual	05 AUG 2009 - 4 AUG 2010

4. TITLE AND SUBTITLE

Engineering Improvements in a Bacterial Therapeutic Delivery System for Breast Cancer

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-08-1-0720

5c. PROGRAM ELEMENT NUMBER

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHOR(S)

Michael McClelland, Ph.D.

mmclelland@vrisd.org

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Vaccine Research Institute of
San Diego
10835 Road to the Cure,
Suite 150
San Diego, CA 92121-1130

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research
and Materiel Command
Fort Detrick, MD 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Serendipitously, the bacterium *Salmonella* accumulates 1000-fold more in tumors than in normal tissue and we have shown that *Salmonella* sometimes cures cancer in animal models. We wished to improve *Salmonella* as a therapeutic system for cancer. In this first funding period there were three objectives and all three were achieved. First, in Aim 1, task 1, we were able to identify over 100 gene mutants of *Salmonella* that are better than any current vaccine strains (safe bacteria) at surviving in tumors but were less virulent than wild type bacteria. These mutants are potentially useful for an improved delivery agent. Second, we had previously identified promoters that were preferentially activated in tumors when compared to spleen. In Aim 2, task 1, we further characterized these regions, which are potentially useful to activate therapeutic genes engineered into the bacterium only when the bacterium is in the tumor. Third, in Aim 2, task 2, we screened promoters *in vitro* for their response to pH (tumors are more acidic than the rest of the body) and anoxia (tumors have lower O2 levels than the rest of the body). A comprehensive survey was performed of all candidate promoters. We also developed a set of bioinformatics programs to help perform, store, analyze, and interpret these experiments, as a foundation for year two of the project.

15. SUBJECT TERMS

High-throughput promoter identification. Genome-wide gene fitness profiles. Bacterial therapeutic delivery system.

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER O206F PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU206	206	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

INTRODUCTION:

A substantial number of breast cancer patients initially present with metastases, or are at substantial risk of relapse after surgery. A therapeutic that would seek out and destroy these metastases, either alone or in combination with other therapies, would be of substantial benefit.

Salmonella enterica sv *Typhimurium*, a facultative anaerobic bacterium that infects both mice and humans, naturally accumulates in a wide variety of solid tumors versus normal mouse tissue at a ratio of 1000:1 (*1*), seemingly preferring the tumor environment over any other niche in the host. The bacterium has been used successfully to selectively kill tumors (*2-4*) and to deliver proteins for cytotoxic or other therapeutic strategies to tumor tissue in mice (*5-16*). We have shown that the *Typhimurium* A1 strain (*leu, arg*) effectively reduces the growth of PC3 and breast tumor xenografts in nude mice while being virtually avirulent in this host organism (5,6). Recently we observed cures of orthotopic human PC3 cancer metastases by *Salmonella* in mice (*17*).

In the first year of this project, we have screened mutations in all non-essential genes in *Salmonella* to identify mutants that are unable to grow well in normal host tissues, and are therefore harmless to humans, but thrive in cancer models *in vivo*. In addition, we have screened for *Salmonella* promoters that are preferentially active in the tumor environment. These promoters can be used to selectively express cloned therapeutic proteins in tumors and export them outside the bacterium, if necessary, while minimizing the side effects of such therapeutics in the rest of the body (*18-20*). Improved growth specificity in tumors combined with expression of therapeutics from promoters with preferential activity in tumor tissues may result in a very specific and inexpensive vector for control of metastases.

Although funding is organized by tissue of origin, it makes scientific sense not to confine our data only to one tissue of origin. Success in attacking breast cancer will be improved if we can demonstrate safety in any cancer type. Similarly, good performance of our Salmonella mutants in any cancer enhances the chances of success in breast. Thus, in a set of experiments of direct relevance to the tasks and goals of this current project, which is designed to test safety and efficacy of one tumor type, we continue to actively pursue the properties of avirulent *Salmonella* in other tumor types and have published a series of papers in this project year that impact on the project (*21-23*). In the first year of the project, two other manuscripts are already in preparation on the tasks in this project.

The approaches we have taken in this project have generated data of a kind never previously generated, or for which tools have never been developed, or both. This fact has required us to develop new analysis tools because such tools did not exist. Our work to develop such tools is a vital and enduring product of this project. Furthermore, all of our tools are made available on the web, making them available to others funded by this mechanism and by other mechanisms at DOD. In the current reporting period we submitted one paper on such a tool, which is widely applicable, not only to Salmonella in breast cancer (Xia et al, WebArrayDB: cross-platform microarray data analysis and public data repository. Submitted. See appendix). We also submitted a methods paper on the use of this tool (Wang et al., Analyzing microarray data using WebArray. Submitted. See appendix). We also submitted a paper on improving oligonucleotide selection for arrays, which also improves the tasks in this project. In the first year of the project, two other manuscripts are already in preparation.

A review article that includes some of the strains and topics in this project was published in the reporting period (*24*).

BODY:

Aim 1. Task 1. Screen for fitness mutants in the tumor and normal tissue environment using a library of transposon-tagged *Salmonella*. (year 1);

In the reporting period a manuscript was planned for the data from this task and will be submitted in the next reporting period. The experiments involved 4T1 breast tumor lines, prostate tumor lines and melanoma (MDA-MB-435) tumors on the theory that the most successful Salmonella strains for therapy would target diverse tumors.

Microarray analysis to determine fitness in normal tissues and tumors. A library of 40,000 *Salmonella* transposon mutants containing mini-Tn5 transposon insertions was injected into twelve tumors growing in 12 nude mice. Three tumor-free mice were injected intravenously with the same *Salmonella* library. Bacteria were recovered after two days from tumors and from the spleens, livers, and lungs of tumor-free mice.

During *in vivo* selection, mutants in genes contributing to fitness in that selective environment are lost from the library. Differences in the mutant library composition before (input library) and after selection (output library) can be detected using microarray hybridization: The transposon sequence carries the T7 promoter sequence, allowing the specific amplification of genomic sequences adjacent to each insertion, which are then mapped on the *Salmonella* genome using a gene microarray. This study revealed two distinct classes of phenotypes: **Class 1 mutants.** This class contains *Salmonella* mutants with reduced fitness in normal tissues (spleen, liver, lung) and unchanged fitness in tumors. We identified mutants affecting at least 19 distinct genes within the SPI-2 island (e.g., *ssrA*, *ssaB*, *ssaC*, *ssaD*, *sseB, sscA, sseC, sseE, ssaJ*, STM1410, *ssaK, ssaL, ssaM, ssaV, ssaN, ssaP, ssaQ, yscR, ssaT*). In addition, mutations in genes involved in a number of cellular functions were identified. These include *htrA, phoP*, and *sifA* and a hypothetical operon containing a putative acetyl-COA hydrolase (STM3118), a putative monoamine oxidase (STM3119) and two putative lysR family transcriptional regulators (STM3120, STM3121). Many of these mutations have previously been observed to be associated with fitness in spleen (25,26). The observation of a similar effect on fitness in liver and lung is new, though not unexpected. The fact that these mutants remain fit in tumors relative to other mutants is new and of potential practical importance for Salmonella use as a direct therapy or for therapy delivery.

Class 2 mutants. This class contains mutants with reduced fitness both in normal tissues and in tumor tissues. Three mutants of the same operon involved in the synthesis of aromatic compounds were identified: *aroM, aroD* and *aroA*. Previous reports describe the use of *Salmonella aroA* and *aroD* mutants in cancer therapy (*27*). Mutants of lipopolysacharide genes belonging to the *rfa* and *rfb* clusters were identified in this class (e.g., *rfbK, rfbM, rfbC, rfaQ*). While class 2 mutants are either known to be avirulent or likely to be of reduced virulence, their impaired growth in tumors relative to class 1 mutants may make them less desirable as strains for delivery of therapy.

Task 2 of this aim in year 2 will take some of these mutants and test them in tumors.

Tumor targeting of STM3120 using syngeneic orthotopic 4T1 breast tumors. The tumor targeting capability of the STM3120 knockout mutant was tested following intragastric delivery into 4T1 murine breast tumor growing orthotopically in the mammary fat pads of BALB/c mice. Five six week-old BALB/c mice bearing 4T1 tumors were each orally injected with 7×10^8 cfu of STM3120, tumor biopsies were taken 2, 5, 7 and 9 days later using Gallini Medical Devices needles and bacterial counts determined. Bacteria were detected in three mice 7 days after administration. At day 9, bacterial counts ranged from 2×10^4 to 9×10^5 cfu per biopsy in all 5 mice.

Table 1. Growth of *STM3120* mutant in orthotopic 4T1 breast tumors after intragastric delivery. Numbers represent cfu per biopsy per mouse taken at different days following injection. Dash indicates a level of bacteria below detection.

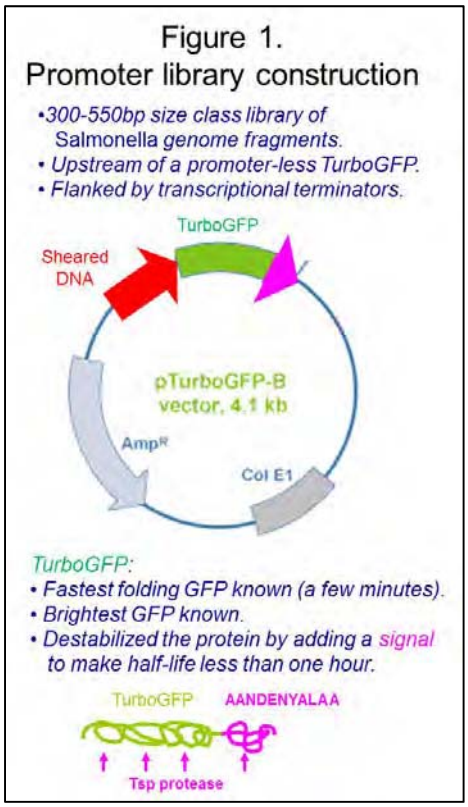
Days \ mouse	1	2	3	4	5
2	-	-	-	-	-
5	-	-	-	-	-
7	-	2.00E+05	1.25E+02	4.00E+05	-
9	3.00E+04	7.50E+05	5.00E+04	9.00E+05	2.00E+04

These results (summarized as Table 1) suggest that intragastric delivery of STM3120 allows a sufficient number of bacteria (~10^7 - 5 x 10^8 cfu per tumor) to target and multiply in the tumor environment to levels that have previously been shown to effectively reduce tumor size after intratumoral or intravenous injections (*30*). This is of importance because intragastric delivery of a therapeutic *Salmonella* strain offers increased convenience over intravenous delivery. A similar finding was recently made by Jia and coworkers (*31*), showing a significant anticancer effect of orally administered VNP20009 into C57b6 mice bearing syngeneic subcutaneous B16F10 melanoma and Lewis lung carcinoma.

Class 1 mutants that retain tumor-targeting while being poor colonizers of normal tissue seem best suited for delivery of cancer therapeutics. However, mutants will need to be tested in the intended host, whether it be humans or companion animals with cancer, before the best candidates for the host can be determined. We have shown that high-throughput transposon library screening allows the identification of novel *Salmonella* mutations of potential therapeutic value, and also allows the re-evaluation of *Salmonella* mutants previously used in cancer therapy. Such approaches can be adapted to any host and tumor model and a wide variety of bacterial species.

Aim 1. Task 2. To test individual mutants for avirulence and tumor-selective properties (year 2 and 3).

In mid-2009 plans were well on the way to screen some of the above mutants in tumors. In addition. Plans were in place to screen over 1000 individual mutants for avirulence in mice. The results of these experiments will be reported in the year two annual report.



Aim 2. Task 1. To identify DNA sequences that act as promoters in tumors but not in normal tissue (year 1).

This step of the project was partly published (32). Work building on this advance has been filed as a patent (attached as an appendix).

Screening of in-vivo tumor-activated promoters. GFP-promoter libraries constructed in a vector that we created (Figure 1) were mixed and injected IT into four human tumor-bearing nude mice. After two days, tumors were combined, homogenized and analyzed by FACS. GFP-positive cells were recovered and expanded overnight in LB containing ampicillin. To eliminate clones harboring constitutive promoters, the tumor library was subjected to a negative FACS sort after overnight growth in LB and a subsequent second positive FACS sort 2 days after a second passage in tumors. We have optimized the FACS analysis to discriminate between true green cells and other fluorescent particles. This was possible by measuring the ratio of fluorescence/auto-fluorescence versus side scatter on the X-axis. **Figure 2** shows the FACS analysis of a sub-library after 2 passages in tumors.

Genome-wide survey of tumor-activated promoters using Nimblegen arrays. Plasmid DNA was extracted from the original promoter library (Library-0), from a sub-library of clones activated in spleen, and from the sub-libraries of clones activated in subcutaneous PC3 tumors in nude mice after two passages in tumors. Promoter sequences were recovered by PCR and labeled by CY5 (Library-0) and CY3 (spleen or tumor library) and then hybridized to an array of 387,000 oligonucleotide sequences spaced at 12 base intervals around the Typhimurium genome (NimbleGen). Using a threshold of two-fold in hybridization signal relative to the control (Library-0), there were 86 intergenic regions enriched in tumor but not in the spleen. Twenty-two intergenic regions are already cloned (see table below) and 174 intergenic regions enriched in both tumor and spleen (data not shown).

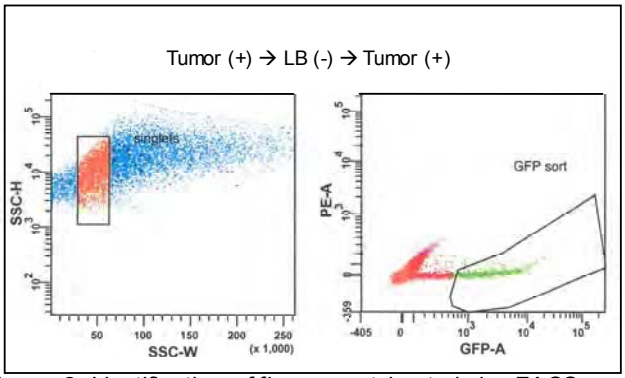


Figure 2. Identification of fluorescent bacteria by FACS

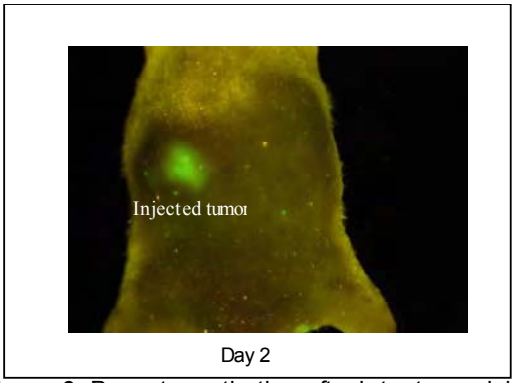


Figure 3. Promoter activation after intra-tumor injection

Table 2: Intergenic regions that induce higher GFP expression in tumor than in spleen

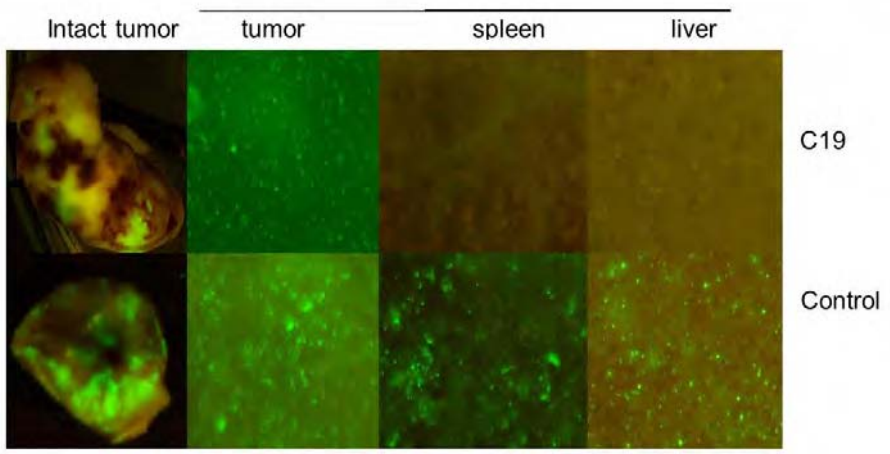
Median of experiment versus input library

Sequenced clone	Spleen	Tumor(+)	Tumor(+)(-)(+)	Tumor(+)(-)(+)	Genes and intergenic regions	5' gene	5' gene orientation	cloned promoter orientation	3' gene	3' gene orientation
	lib1	lib2	lib3	lib4						
	Sequenced clones:									
85	0.9	2.3	5.5	9.5	IR STM0468 - STM0469	ylaB	-	+	rpmE2	+
86	1.9	1.7	3.2	2.6	IR STM0474 - STM0475	ybaJ	-	-	acrB	-
	1.3	1.6	3.5	3.0	STM0475					
	1.1	3.7	0.3	8.2	STM0580					
87	0.9	3.2	0.3	8.5	IR STM0580 - STM0581	STM0580	-	-	STM0581	+
	0.8	2.6	3.8	0.4	STM0844					
10	2.9	1.9	8.9	0.3	IR STM0844 - STM0845	pflE	-	-	moeB	-
	0.8	0.7	5.8	0.4	STM0845					
	0.7	3.5	5.4	7.0	STM0937					
11	0.7	4.2	6.5	###	IR STM0937 - STM0938	hcp	-	-	ybjE	-
	0.5	3.9	7.1	###	STM0938					
	0.7	6.5	9.0	###	STM1382					
16	0.7	4.6	7.4	###	IR STM1382 - STM1383	orf408	-	-	ttrA	-
20	1.9	5.5	2.8	###	IR STM1529 - STM1530	STM1529	+	-	STM1530	+
	0.8	2.5	6.3	###	STM1807					
26	1.1	1.6	6.6	###	IR STM1807 - STM1808	dsbB	+	+	STM1808	+
	0.6	2.3	4.1	###	STM1808					
	0.9	3.2	5.4	9.0	STM1914					
28	0.9	3.9	7.2	7.5	IR STM1914 - STM1915	flhB	-	-	cheZ	-
	1.2	4.1	5.8	7.3	STM1915					
	1.2	2.9	6.6	3.5	STM1996					
30	1.2	2.9	7.4	4.0	IR STM1996 - STM1997	cspB	-	-	umuC	-
31	1.3	5.9	4.7	7.9	IR STM2035 - STM2036	cbiA	-	-	pocR	-
34	0.6	2.1	3.5	4.7	IR STM2261 - STM2262	napF	-	-	eco	+
	0.7	3.5	6.3	7.0	STM2309					
36	0.6	2.7	6.5	6.3	IR STM2309 - STM2310	menD	-	-	menF	-
	0.6	3.3	6.5	6.1	STM2310					
	1.0	1.1	2.1	2.4	STM3070					
44	0.8	1.4	2.8	3.1	IR STM3070 - STM3071	epd	-	-	STM3071	+
	1.1	2.6	3.9	4.5	STM3106					
45	1.1	3.5	4.6	4.6	IR STM3106 - STM3107	ansB	-	-	yggN	-
	1.2	2.4	4.4	3.1	STM3107					
55	0.8	3.8	1.8	5.6	IR STM3525 - STM3526	glpE	-	+	glpD	+
	1.0	4.2	2.2	6.9	STM3526					
	0.8	5.6	6.2	###	STM3880					
61	0.9	5.4	6.1	###	IR STM3880 - STM3881	kup	+	+	rbsD	+
	0.6	5.6	2.8	###	STM3881					
	1.0	2.4	3.2	6.1	STM4289					
71	0.9	2.0	8.3	9.6	IR STM4289 - STM4290	phnA	-	-	proP	+
77	0.8	3.4	8.3	6.0	IR STM4418 - STM4419	STM4418	-	+	STM4419	+
	0.8	3.4	7.1	3.8	STM4419					
78	1.3	6.1	5.6	8.0	IR STM4430 - STM4431	STM4430	-	+	STM4431	+
	0.8	2.2	5.6	7.4	STM4431					

Sequencing of Promoters. 192 clones from the tumor library were picked at random, sequenced, and 100 different sequences were obtained. These sequences were mapped to the genome and their potential regulation (tumor-specific or active in both spleen and tumor) was determined by comparison with the microarray data. We found 22 candidate promoters preferentially activated in tumors and 40 candidates constitutive promoters. Tumor-specific clones recovered in this experiment represent 23% of the total 95 tumor-expressed intergenic regions detected on microarrays. **Table 2** includes promoter fragments that were cloned that showed differential activity on the array assay.

Confirmation of tumor specificity of individual clones *in vivo*. Twenty-two tumor-specific candidates were recovered; of these three were individually confirmed *in vivo*. The clones were intravenously injected at $5x10^6$, $1x10^7$ and $5x10^7$ cfu into tumor-free and tumor-bearing nude mice. One or two days post-injection, spleens and tumors were imaged using the OV100, homogenized, and the bacterial titer was quantified on LB+Amp plates. Spleens from normal mice were compared with tumors that had similar bacteria counts, so that any difference in fluorescence would be attributable to increased GFP expression rather than bacterial numbers. **Figure 4** presents images that indicate that the tumors are much more fluorescent than spleens infected with the same number of bacteria for each of the three clones. Contrary to these putative tumor-specific clones, a positive control that constitutively expresses TurboGFP resulted in strong fluorescence in spleen even with doses as low as $2x10^5$ cfu. An example is shown in figure 4 for promoter clone C19.

Figure 4. GFP-based promoter expression in tumors and normal tissues in nude mice using the whole mouse OV100 imaging system. Promoter clone C19 is expressed in tumors (GFP positive) but not in spleens (GFP negative), a constitutive GFP promoter pturbo (control), is activated in both tissues.



Regulatory pathways for promoters preferentially induced in tumors. Promoters regulated by anaerobiosis are likely to be induced in the hypoxic regions of solid tumors and most of them are under control of the *Salmonella* global regulators Fnr and ArcA (*1*). There are at least 22 candidate promoters of this class among the 95 tumor-specific intergenic regions identified on arrays (data not shown); two of the anaerobic induced promoters are shown in Figure 4. **Clone 10** is the promoter region of a putative pyruvate formate lyase activating enzyme (pflE) and the promoter region of pflE contains Fnr regulated sequence. In *E. coli,* the anaerobic transcription of the next gene (pflF) is co-regulated by two major global regulators of anaerobic metabolism, ArcA and Fnr (*1*). **Clone 45** contains the promoter region of *ansB* which encodes part of asparaginase, a tetrameric enzyme that catalyzes the hydrolysis of asparagine to aspartic acid and ammonia. In *E. coli*, *ansB* is positively co-regulated by CRP (cyclic AMP receptor protein) and the Fnr protein (*1*). However, in *Salmonella enterica* the anaerobic regulation of the *ansB* gene may require only CRP (*1*). **Clone 28** contains the promoter region of *flhB*, a gene that is required for the formation of the rod structure of the flagellar apparatus (*1*). This candidate promoter and many others identified on arrays are not known to be induced by hypoxia. Some of these promoters may be induced by a different signal present in subcutaneous tumors.

Transition to the second year. In aim 2, task 2, below, we will discuss further improvements in the approach and the development of tools for those approaches. We also want to repeat experiments in orthotopic models.

Aim 2. Task 2. To identify promoters that respond to anoxia and/or acid pH, or neither (year 1).

We subjected our Salmonella strain to anoxia and to various pHs and grew them to stationary phase.. This was done in triplicate. Then the samples were subjected to FACS sorting. The resulting material was applied to Nimblegen arrays (tiling arrays of 387,000 overlapping oligos in both strands of the Salmonella genome). No tools existed for analyzing this kind of data and indeed, no tools other than ours have been developed since that time. So we embarked on developing those tools. At the end of the first year, these tools were still in raw form. The accompanying figure shows the power of the data using one type of presentation from the first tool developed. Tabular calculations identified tens of differentially induced promoters which will be the topic of the next annual report. At least ten promoters with responsiveness to pH and anoxia will be under investigation for year 2. Of special interest are those promoters that were found in surveys of tumor versus spleen (Task 1), and particularly if these are under different regulation in vitro. Combinations of such promoters could lead to tight regulation of therapeutics only in anoxia + low pH, for example.

Figure 5 shows some of visualization efforts. Note that promoter data has never been presented at this resolution, and this comprehensive, and in this user-friendly form before. In future reporting periods we hope to convert animal promoter experiments to the same analysis pipeline.

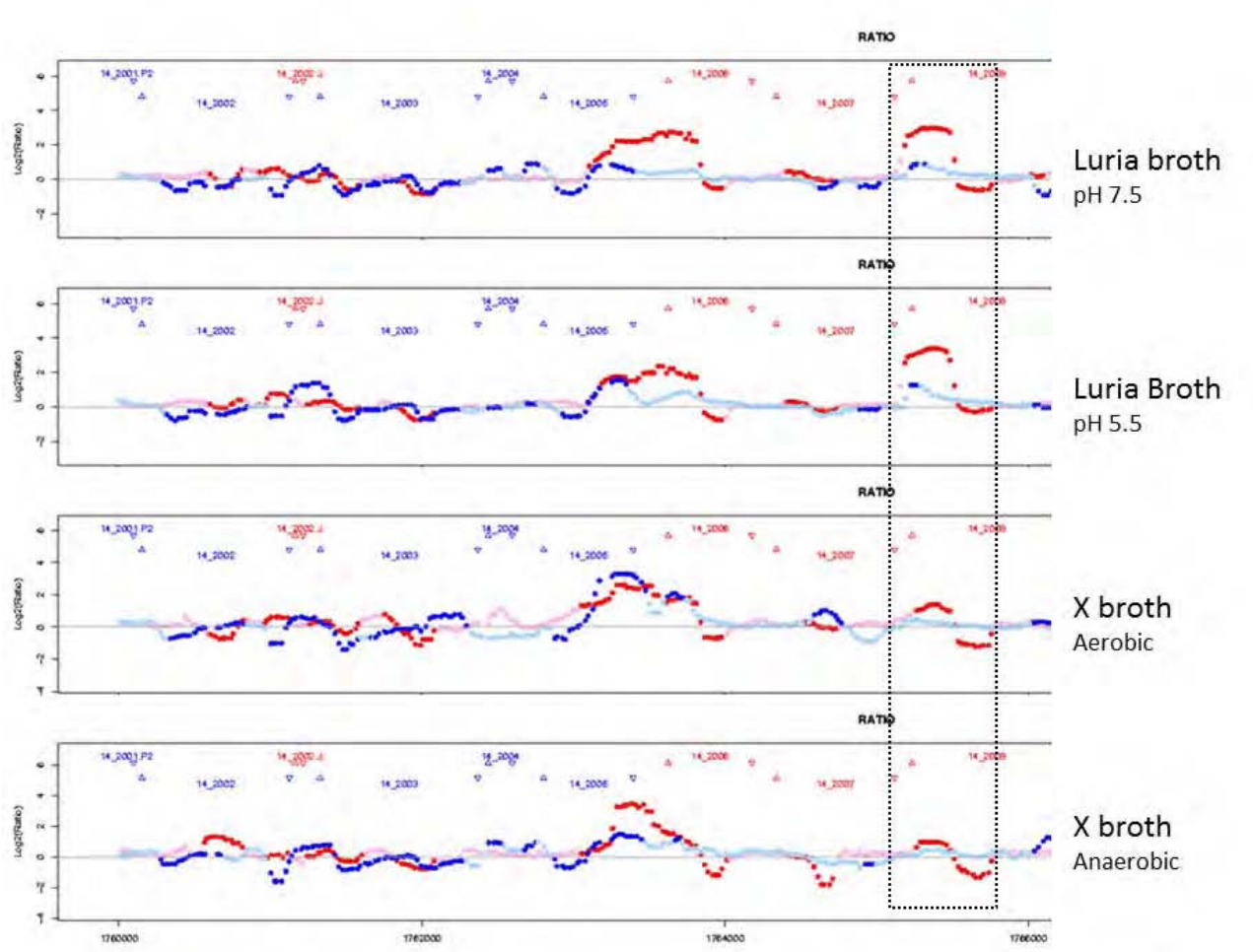


Figure 5. A comprehensive survey indicates differentially activated promoters. This example is a promoter that differs between growth conditions. In this figure 0.2% of the entire genome is presented in the X axis. Blue indicates genes in the sense strand and red in the antisense strand. Gene starts are upright and inverted triangles. The strand of the captured promoter is also presented in the same colors. The Y axis represents the log2 of the ratio of the input library to the FACS sorted library. The region indicated by a box shows a promoter that is four-fold to eight-fold more active in Luria broth than in X-broth (a media often used for anaerobic growth).

Development of new tools for the accomplishment of the tasks in this project. The analysis of the data in Aim 2, tasks 1 and 2 required a considerable amount of bioinformatics development. Mapping of promoters to microarrays is a non-trivial task. The work for the figure above was not ready for publication as a tool, but many intermediate steps were ready. This work is described in the attached manuscripts; Xia et al, WebArrayDB: cross-platform microarray data analysis and public data repository. Submitted. See appendix; and in Wang et al., Analyzing microarray data using WebArray. Submitted. See appendix. In brief, an open source integrated microarray database and analysis suite, WebArrayDB (http://www.webarraydb.org), was developed that features convenient uploading of data for storage in a MIAME (Minimal Information about a Microarray Experiment) compliant fashion, and allows data to be mined with a large variety of R-based tools, including data analysis across multiple platforms. Different methods for probe alignment, normalization and statistical analysis were included to account for systematic bias. Student's t-test, moderated t-tests, non-parametric tests and analysis of variance or covariance (ANOVA/ANCOVA) are among the choices of algorithms for differential analysis of data. Users also have the flexibility to define new factors and create new

analysis models to fit complex experimental designs. All data can be queried or browsed through a web browser. The computations can be performed in parallel on symmetric multiprocessing (SMP) systems or Linux clusters. The software package is available for the use on a public web server (http://www.webarraydb.org) or can be downloaded at Bioinformatics online.

We have spent considerable effort to improve oligo selection for arrays (Xia et al., Evaluating oligonucleotide properties for DNA microarray probe design. Submitted, see Appendix). In brief, Most current microarray oligonucleotide probe design strategies are based on probe design factors (PDFs), that include probe hybridization free energy (PHFE), probe minimum folding energy (PMFE), dimer score, hairpin score, homology score, and complexity score. The impact of these PDFs on probe performance was evaluated using four sets of microarray comparative genome hybridization (aCGH) data, which included two array manufacturing methods and the genomes of two species; Salmonella and humans. We developed a new probe design factor, pseudo probe binding energy (PPBE), by iteratively fitting di-nucleotide positional weights and di-nucleotide stacking energies until the average residue sum of squares (ARSS) for the model was minimized. PPBE showed a better correlation with probe sensitivity and a better specificity than all other PDFs. The physical properties that are measured by PPBE are as yet unknown but include a platform-dependent component. Programs and correlation parameters from this study are freely available to facilitate the design of DNA microarray oligonucleotide probes.

Using these tools we are able to generate, store, analyze, and present data in an expeditious, useful, and attractive manner.

Aim 2. Task 3. To test individual candidate promoters for differential activity in tumors (year 2).

In mid 2009, plans were in place to screen candidate promoters and this task for year two will be reported in the year two annual report.

Aim 3. To combine the best mutant strains with the best tumor-specific promoters (year 2).

In mid 2009, plans were in place to screen candidate promoters and this task for year two will be reported in the year two annual report.

Aim 4. To test one potential therapeutic delivery system as a proof of principle (as time permits).

Although this step was beyond the scope of the proposal and was not a formal task, we have signed MTAs and acquired the vectors needed for this step. It is not likely we will perform this optional task until year 3, at the earliest.

KEY RESEARCH ACCOMPLISHMENTS:

- Identification of a few candidate genes that when mutated alter the targeting of *Salmonella* to cancers.
- Identification of over 50 candidate *Salmonella* genes that when mutated allow growth in tumors while debilitating virulent growth in the spleen.
- Identification of over 50 candidate *Salmonella* promoter regions that are preferentially activated in tumors but not in the spleen.
- Identification of tens of promoters in vitro responsive to anoxia and pH that potentially correlate with conditions in the tumor.

- Improvements in data analysis software with the continued updating of www.webarrayDB, a public resource that we maintain so that it can handle new kinds of data.

REPORTABLE OUTCOMES:

Abstracts presented:

Nabil Arrach, Ming Zhao, Steffen Porwollik, Robert M. Hoffman and Michael McClelland (2008) *Salmonella* promoters preferentially activated inside tumors. Annual meeting of the American Association for Cancer Research, San Diego, California, USA

Nabil Arrach, Ming Zhao, Robert M. Hoffman and Michael McClelland (2009) Microarray screening of *Salmonella* variants for tumor targeting. Annual meeting of the American Association for Cancer Research, Denver, Colorado, USA

Submitted:

Xia XQ, Jia Z, Porwollik S, Long F, Hoemme C, Ye K, Müller-Tidow C, McClelland M, Wang Y. Evaluating oligonucleotide properties for DNA microarray probe design.

Xia XQ, McClelland M, Porwollik S, Song W, Cong X, Wang Y. WebArrayDB: cross-platform microarray data analysis and public data repository.

Wang Y, McClelland M, Xia XQ. Analyzing Microarray Data Using WebArray.

Planned or in preparation:

Santiviago CA, Reynolds MM, Porwollik S, Choi SH, Long F, Andrews-Polymenis HL, McClelland M. Analysis of pools of targeted *Salmonella* deletion mutants identifies novel genes affecting fitness during competitive infection in mice.

Wang Y, Xia XQ, Zhenyu Jia Z, Anne Sawyers A, Yao H, Wang-Rodriquez J, Mercola D, McClelland M. *In silico* estimates of tissue components in surgical samples based on expression profiling data.

Xia XQ, McClelland M, Wang Y. TabSQL: a MySQL tool to facilitate mapping user data to public databases.

Arrach N, Cheng P, Zhao M, Santiviago CA, Hoffman RM, McClelland M. High-throughput screening for Salmonella avirulent mutants that retain targeting of solid tumors.

Patent applications filed:

PCT VIV-1001-PC, Methods to treat solid tumors (see appendix).

Informatics and databases:

Improvements to www.webarrayDB.org and to oligo selection methods for arrays.

Improvements to databasing; increased capacity and ease of use.

CONCLUSION:

In Aim 1, task 1, the investigators have identified over 50 genes that share the desirable feature of rendering Salmonella less virulent for infection but which still retain the ability to target tumors and grow in tumors. The ability to target tumors after oral delivery was also demonstrated. This sets the stage for Aim 1, Task 2, in years 2 and 3, in which we will test individual mutants for avirulence and tumor-selective properties.

In Aim 2, task 1, over 50 candidate promoters were identified that were induced in tumor, and may be less induced in other parts of the animal host. In Aim 2, task 2, all the experiments on anoxia and pH were completed. A few genes induced preferentially in these conditions are under investigation. Tools were developed to present this data and have been made publicly available. This sets the stage for Aim 2, Task 3, in years 2 and 3, in which we will test individual candidate promoters for differential activity in tumors.

REFERENCES:

1. Pawelek, J.M., Low, K.B. and Bermudes, D. (1997) Tumor-targeted Salmonella as a novel anticancer vector. *Cancer research*, 57, 4537-4544.

2. Saltzman, D.A. (2005) Cancer immunotherapy based on the killing of Salmonella typhimurium-infected tumour cells. *Expert Opin Biol Ther*, 5, 443-449.

3. Barnett, S.J., Soto, L.J., 3rd, Sorenson, B.S., Nelson, B.W., Leonard, A.S. and Saltzman, D.A. (2005) Attenuated Salmonella typhimurium invades and decreases tumor burden in neuroblastoma. *J Pediatr Surg*, 40, 993-997; discussion 997-998.

4. Lee, C.H., Wu, C.L., Tai, Y.S. and Shiau, A.L. (2005) Systemic administration of attenuated Salmonella choleraesuis in combination with cisplatin for cancer therapy. *Mol Ther*, 11, 707-716.

5. Zhao, M., Yang, M., Li, X.M., Jiang, P., Baranov, E., Li, S., Xu, M., Penman, S. and Hoffman, R.M. (2005) Tumor-targeting bacterial therapy with amino acid auxotrophs of GFP-expressing Salmonella typhimurium. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 755-760.

6. Zhao, M., Yang, M., Ma, H., Li, X., Tan, X., Li, S., Yang, Z. and Hoffman, R.M. (2006) Targeted therapy with a Salmonella typhimurium leucine-arginine auxotroph cures orthotopic human breast tumors in nude mice. *Cancer research*, 66, 7647-7652.

7. Clairmont, C., Lee, K.C., Pike, J., Ittensohn, M., Low, K.B., Pawelek, J., Bermudes, D., Brecher, S.M., Margitich, D., Turnier, J. *et al.* (2000) Biodistribution and genetic stability of the novel antitumor agent VNP20009, a genetically modified strain of Salmonella typhimurium. *J Infect Dis*, 181, 1996-2002.

8. Nishikawa, H., Sato, E., Briones, G., Chen, L.M., Matsuo, M., Nagata, Y., Ritter, G., Jager, E., Nomura, H., Kondo, S. *et al.* (2006) In vivo antigen delivery by a Salmonella typhimurium type III secretion system for therapeutic cancer vaccines. *J Clin Invest*, 116, 1946-1954.

9. Panthel, K., Meinel, K.M., Sevil Domenech, V.E., Geginat, G., Linkemann, K., Busch, D.H. and Russmann, H. (2006) Prophylactic anti-tumor immunity against a murine fibrosarcoma triggered by the Salmonella type III secretion system. *Microbes and infection / Institut Pasteur*, 8, 2539-2546.

10. Avogadri, F., Martinoli, C., Petrovska, L., Chiodoni, C., Transidico, P., Bronte, V., Longhi, R., Colombo, M.P., Dougan, G. and Rescigno, M. (2005) Cancer immunotherapy based on killing of Salmonella-infected tumor cells. *Cancer research*, 65, 3920-3927.

11. Thamm, D.H., Kurzman, I.D., King, I., Li, Z., Sznol, M., Dubielzig, R.R., Vail, D.M. and MacEwen, E.G. (2005) Systemic administration of an attenuated, tumor-targeting Salmonella typhimurium to dogs with spontaneous neoplasia: phase I evaluation. *Clin Cancer Res*, 11, 4827-4834.

12. Forbes, N.S., Munn, L.L., Fukumura, D. and Jain, R.K. (2003) Sparse initial entrapment of systemically injected Salmonella typhimurium leads to heterogeneous accumulation within tumors. *Cancer research*, 63, 5188-5193.

13. Meng, F.P., Ding, J., Yu, Z.C., Han, Q.L., Guo, C.C., Liu, N. and Fan, D.M. (2005) Oral attenuated Salmonella typhimurium vaccine against MG7-Ag mimotope of gastric cancer. *World J Gastroenterol*, 11, 1833-1836.

14. Nemunaitis, J., Cunningham, C., Senzer, N., Kuhn, J., Cramm, J., Litz, C., Cavagnolo, R., Cahill, A., Clairmont, C. and Sznol, M. (2003) Pilot trial of genetically modified, attenuated Salmonella expressing the E. coli cytosine deaminase gene in refractory cancer patients. *Cancer Gene Ther*, 10, 737-744.

15. Bermudes, D., Low, K.B., Pawelek, J., Feng, M., Belcourt, M., Zheng, L.M. and King, I. (2001) Tumour-selective Salmonella-based cancer therapy. *Biotechnol Genet Eng Rev*, 18, 219-233.

16. Tjuvajev, J., Blasberg, R., Luo, X., Zheng, L.M., King, I. and Bermudes, D. (2001) Salmonella-based tumor-targeted cancer therapy: tumor amplified protein expression therapy (TAPET) for diagnostic imaging. *J Control Release*, 74, 313-315.

17. Zhao, M., Geller, J., Ma, H., Yang, M., Penman, S. and Hoffman, R.M. (2007) Monotherapy with a tumor-targeting mutant of Salmonella typhimurium cures orthotopic metastatic mouse models of human prostate cancer. *Proceedings of the National Academy of Sciences of the United States of America*.

18. Weyel, D., Sedlacek, H.H., Muller, R. and Brusselbach, S. (2000) Secreted human beta-glucuronidase: a novel tool for gene-directed enzyme prodrug therapy. *Gene Ther*, 7, 224-231.

19. Lee, S.H. and Camilli, A. (2000) Novel approaches to monitor bacterial gene expression in infected tissue and host. *Curr Opin Microbiol*, 3, 97-101.

20. Galen, J.E., Zhao, L., Chinchilla, M., Wang, J.Y., Pasetti, M.F., Green, J. and Levine, M.M. (2004) Adaptation of the endogenous Salmonella enterica serovar Typhi clyA-encoded hemolysin for antigen export enhances the immunogenicity of anthrax protective antigen domain 4 expressed by the attenuated live-vector vaccine strain CVD 908-htrA. *Infection and immunity*, 72, 7096-7106.

21. Nagakura, C., Hayashi, K., Zhao, M., Yamauchi, K., Yamamoto, N., Tsuchiya, H., Tomita, K., Bouvet, M. and Hoffman, R.M. (2009) Efficacy of a genetically-modified Salmonella typhimurium in an orthotopic human pancreatic cancer in nude mice. *Anticancer research*, 29, 1873-1878.

22. Hayashi, K., Zhao, M., Yamauchi, K., Yamamoto, N., Tsuchiya, H., Tomita, K. and Hoffman, R.M. (2009) Cancer metastasis directly eradicated by targeted therapy with a modified Salmonella typhimurium. *Journal of cellular biochemistry*, 106, 992-998.

23. Hayashi, K., Zhao, M., Yamauchi, K., Yamamoto, N., Tsuchiya, H., Tomita, K., Kishimoto, H., Bouvet, M. and Hoffman, R.M. (2009) Systemic targeting of primary bone tumor and lung metastasis of high-grade osteosarcoma in nude mice with a tumor-selective strain of Salmonella typhimurium. *Cell cycle (Georgetown, Tex*, 8, 870-875.

24. Hoffman, R.M. (2009) Tumor-targeting amino acid auxotrophic Salmonella typhimurium. *Amino acids*, 37, 509-521.

25. Santiviago, C.A., Reynolds, M.M., Porwollik, S., Choi, S.H., Long, F., Andrews-Polymenis, H.L. and McClelland, M. (2009) Analysis of pools of targeted Salmonella deletion mutants identifies novel genes affecting fitness during competitive infection in mice. *PLoS pathogens*, 5, e1000477.

26. Chan, K., Kim, C.C. and Falkow, S. (2005) Microarray-based detection of Salmonella enterica serovar Typhimurium transposon mutants that cannot survive in macrophages and mice. *Infection and immunity*, 73, 5438-5449.

27. Lawley, T.D., Chan, K., Thompson, L.J., Kim, C.C., Govoni, G.R. and Monack, D.M. (2006) Genome-wide screen for salmonella genes required for long-term systemic infection of the mouse. *PLoS pathogens*, 2, e11.

28. Datsenko, K.A. and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America*, 97, 6640-6645.

29. Yoon, W.S., Choi, W.C., Sin, J.I. and Park, Y.K. (2007) Antitumor therapeutic effects of Salmonella typhimurium containing Flt3 Ligand expression plasmids in melanoma-bearing mouse. *Biotechnology letters*, 29, 511-516.

30. Bermudes, D., Zheng, L.M. and King, I.C. (2002) Live bacteria as anticancer agents and tumor-selective protein delivery vectors. *Curr Opin Drug Discov Devel*, 5, 194-199.

31. Jia, L.J., Wei, D.P., Sun, Q.M., Huang, Y., Wu, Q. and Hua, Z.C. (2007) Oral delivery of tumor-targeting Salmonella for cancer therapy in murine tumor models. *Cancer science*, 98, 1107-1112.

32. Arrach, N., Zhao, M., Porwollik, S., Hoffman, R.M. and McClelland, M. (2008) Salmonella promoters preferentially activated inside tumors. *Cancer research*, 68, 4827-4832.

APPENDICES:

A patent application filed for worldwide protection under the Patent Cooperation Treaty (PCT).

Xia XQ, Jia Z, Porwollik S, Long F, Hoemme C, Ye K, Müller-Tidow C, McClelland M, Wang Y. Evaluating oligonucleotide properties for DNA microarray probe design. Submitted 2009.

Xia XQ, McClelland M, Porwollik S, Song W, Cong X, Wang Y. WebArrayDB: cross-platform microarray data analysis and public data repository. Submitted 2009.

Wang Y, McClelland M, Xia XQ. Analyzing Microarray Data Using WebArray. Submitted 2009.

GENOMIC TECHNOLOGIES AND APPLICATIONS -- POSTER PRESENTATIONS - PROFFERED ABSTRACTS:
Nabil Arrach, Ming Zhao, Robert Hoffman, and Michael McClelland
Abstract #5210: Microarray screening of Salmonella variants for tumor targeting
AACR Meeting Abstracts, Apr 2009; 2009: 5210.

100th AACR Annual Meeting-- Apr 18-22, 2009; Denver, CO

Genomic Technologies and Applications -- Poster Presentations - Proffered Abstracts

Abstract #5210: Microarray screening of Salmonella variants for tumor targeting

Nabil Arrach, Ming Zhao, Robert Hoffman and Michael McClelland

Sidney Kimmel Cancer Center, San Dieo, CA; Anticancer Inc, San Dieo, CA

Abstract

Salmonella typhimurium has the ability to target a wide range of solid tumors and accumulates thousands of folds in tumors when compared to normal tissues. Only a handful of attenuated Salmonella strains are currently being investigated for cytokine delivery or gene directed enzyme pro-drug therapy. There remains considerable scope for engineering low toxicity and improved targeting to tumors in humans. A high throughput screening of a complex Salmonella mutant library was performed in human prostate tumors, melanomas, and normal tissues in nude mice. Microarrays were used to identify Salmonella variants that have reduced fitness in normal tissues (for safety) but still thrive in tumors (unchanged fitness or even increased fitness). Our data reveal that some Salmonella mutants previously used for cancer therapy, such as aroA and aroD are very safe, but at a disadvantage for growth in tumors. Screening for optimized safe strains can be applied to multiple animal models to ensure the generality of the findings, potentially improving safety and targeting of cancers in humans.

Citation Information: In: Proc Am Assoc Cancer Res; 2009 Apr 18-22; Denver, CO. Philadelphia (PA): AACR; 2009. Abstract nr 5210.

GENE THERAPY 1: POSTER PRESENTATIONS - PROFFERED ABSTRACTS:
Nabil Arrach, Ming Zhao, Steffen Porwollik, Robert Hoffman, and Michael McClelland
Salmonella promoters preferentially activated inside tumors
AACR Meeting Abstracts, Apr 2008; 2008: 1363.

Gene Therapy 1: Poster Presentations - Proffered Abstracts

Abstract #1363

Salmonella promoters preferentially activated inside tumors

Nabil Arrach, Ming Zhao, Steffen Porwollik, Robert Hoffman and Michael McClelland

Sidney Kimmel Cancer Ctr., San Diego, CA, AntiCancer, Inc., San Diego, CA

Salmonella has the ability to preferentially grow in the hypoxic environment of solid tumors and has previously been used to express therapeutic proteins. We have recently developed a strain of *S. typhimurium* which preferentially targets viable tumor tissue as well as necrotic tissue (Proc. Natl. Acad. Sci. USA **104**, 10170-10174, 2007). However, bacteria still circulate at low levels in the body. Control of protein expression usingendogenous *Salmonella* promoters that are preferentially activated in tumors could further improve targeting of therapies. A random library of*Salmonella enterica* Typhimurium 14028 genomic DNA was cloned upstream of a promoter-less green fluorescent protein gene (TurboGFP) and intravenously injected into tumor-free mice and into human PC3 prostate tumors growing subcutaneously in nude mice. After two days, fluorescence-activated cell sorting was used to enrich for bacterial clones expressing GFP in spleens or in tumors. The resulting libraries were hybridized to anoligonucleotide tiling array of the *Salmonella* genome. 95 intergenic regions were enriched in tumor samples but not in spleen. Sequencing of 100 clones from a tumor-enriched library yielded 22 from intergenic regions that showed significant enrichment in tumors versus spleen in the microarrays. Three of these 22 candidate promoter clones were tested *in vivo* and enhanced GFP expression in tumor relative to spleen was confirmed. Two of the three clones mapped to the pflE and ansB promoter regions, which are known to undergo induction in the hypoxic conditions that occur in solid tumors. Most of the other 93 candidates are not known to be regulated by hypoxia and some may reveal other properties of tumors exploited by*Salmonella*. The expression of therapeutics in *Salmonella* under the regulation of one or more promoters that are activated preferentially in tumors has the potential for tumor-targeted therapy with reduced side-effects.

WebArrayDB: Cross-platform microarray data analysis and public data repository

Xiao-Qin Xia [1*], Michael McClelland [1*], Steffen Porwollik [1], Wenzhi Song [1,2], Xianling Cong [3] and Yipeng Wang [1,4*]

[1] *Vaccine Research Institute San Diego, 10835 Road to the Cure, San Diego, CA 92121, USA*

[2] *Department of Oralogy, China Japan Union Hospital, Jilin University, Changchun, 130031, China*

[3] *Department of Dermatology, China Japan Union Hospital, Jilin University, Changchun, 130031, China*

[4] *Department of Pathology & Laboratory Medicine, University of California, Irvine, CA 92697, USA*

Abstract

Cross-platform microarray analysis is an increasingly important research tool, but researchers still lack open source tools for storing, integrating, and analyzing large amounts of microarray data obtained from different array platforms. An open source integrated microarray database and analysis suite, WebArrayDB (http://www.webarraydb.org), has been developed that features convenient uploading of data for storage in a MIAME (Minimal Information about a Microarray Experiment) compliant fashion, and allows data to be mined with a large variety of R-based tools, including data analysis across multiple platforms. Different methods for probe alignment, normalization and statistical analysis are included to account for systematic bias. Student's t-test, moderated t-tests, non-parametric tests, and analysis of variance or covariance (ANOVA/ANCOVA) are among the choices of algorithms for differential analysis of data. Users also have the flexibility to define new factors and create new analysis models to fit complex experimental designs. All data can be queried or browsed through a web browser. The computations can be performed in parallel on symmetric multiprocessing (SMP) systems or Linux clusters.

[WebArrayDB is freely available at http://www.webarraydb.org.]

Introduction

Large amounts of microarray experimental data are stored in public repositories, making cross-platform analysis of data from different sources (either different laboratories and/or different platforms) an increasingly attractive and important research tool [Moreau et al., 2003]. Such analyses are possible because biological treatments usually have a greater impact on measured expression than the noise of a cross-platform analysis [Chen et al., 2008, Larkin et al., 2005, Shippy et al., 2004]. Moreover, the combined use of multiple platforms can overcome the inherent biases of individual platforms for identification of the more robust changes in gene expression profiles [Bosotti et al., 2007].

Currently available analysis packages do not provide all the required functions for cross-platform integration, normalization, and statistical analysis of data from different sources. Integrative Array Analyzer (iArray) [Pan et al., 2006] offers statistical cross-platform analysis functions but does not have probe alignment or data normalization features. MatchMiner [Bussey et al., 2003] is a powerful tool for matching genes and gene products from two platforms but is not designed for statistical analysis. The Gene Expression Pattern Analysis Suite (GEPAS) [Tárraga et al., 2008] integrates many tools for microarray data analysis, but it does not have data storage capability or cross-platform analysis functions. Other online platforms and public repositories are designed mainly for data storage and lack probe matching and cross-platform analysis functions: prominent examples include Expression Profiler [Kapushesky et al., 2004], ArrayExpress [Parkinson et al., 2007], the Stanford Microarray Database (SMD) [Demeter et al., 2007], the Longhorn Array Database (LAD) [Killion et al., 2003] and the BioArray Software Environment (BASE) [Saal et al., 2002, Troein et al.,

*Corresponding authors

2006].

An earlier open-source online platform for microarray data analysis, WebArray [Xia et al., 2005], did not offer a cross-platform analysis function, but provided an excellent framework for extension to WebArrayDB (http://www.webarraydb.org) - a database system and analysis suite that provides this function. In addition to traditional methods such as median and quantile for between-array normalization, WebArrayDB has integrated median rank scores (MRS), quantile discretization (QD) [Warnat et al., 2005], gene quantile (GQ) - a quantile normalization for each individual gene among different platforms, and principal component analysis (PCA) [Stoyanova et al., 2004]. WebArrayDB provides standard statistical analysis methods, such as Student's t-test, eBayes-moderated t-test, Significance Analysis of Microarrays (SAM) [Tusher et al., 2001], ANOVA/ANCOVA and non-parametric tests, as options for users to explore.

Database Infrastructure

WebArrayDB includes all fields required for MIAME-compliant microarray data storage [Brazma et al., 2001]. Data are classified into five categories: "project", "array", "platform", "protocol" and "sample". Each record in these tables is given a unique ID ("MPMDB ID"), and all five categories have to be filled for MIAME compliance and subsequent data analysis. All tables in the database have been indexed to speed up queries even when the size of the data set becomes very large.

The project table serves as the hub of information - most information is linked to a specific project in the database (**Figure 1** and **Supplementary Figure 1**). Intrinsic relationships among project, array, platform, protocol, and sample are directly linked by references between tables, which permits fast cross-table searching. When defining a platform, users may supply probe information, including user-defined IDs and gene IDs from other public databases, such as RefSeq, UniGene, etc. All of these IDs can serve as references for cross-platform probe alignment. Since there are extensive gene annotations in GO (Gene Ontology database, http://www.geneontology.org/) [Ashburner and Lewis, 2002] , WebArrayDB is also designed to facilitate the use of GO for probe searching. The GO database in WebArrayDB is updated monthly.

The project table is linked to the "users" table that contains the user information including user name and password (**Figure 1**), enabling data access to be controlled based on user privileges. Every project has an associated release date which determines the public accessibility of the project. By default the project release date will be two years from the data deposit date to protect data privacy. The user can change the release date at the time the data is deposited or at any time thereafter.

WebArrayDB is powered by the affy [Gautier et al., 2004] and the Linear Models for Microarray Data (LIMMA, http://bioinf.wehi.edu.au/limma) [Smyth, 2005] packages from bioconductor (http://www.bioconductor.org/), which are open source and open development software projects for the analysis and comprehension of genomic data. Thus, many different formats of intensity files are recognized, including data from Affymetrix CEL files, Agilent Feature Extraction, ArrayVision, BlueFuse, GenePix, QuantArray (Version 3 or later), SMD and SPOT. Any formats that affy and LIMMA do not recognize can be accepted when defined by the user in a tab-delimited text file, including data with more than 2 scanned channels.

WebArrayDB stores parsed data in database tables. The image files, intensity files, probe files, protocol files and other user-supplied raw data files are stored in the file system on servers with indices in the database.

Data Analysis

Data queried from the database can be directly subjected to analysis. WebArrayDB presents a variety of options for data preprocessing, and differential analysis. Conservative default analysis methods and parameters are set so that novice users will be less likely to use flawed analysis strategies.

Data preprocessing

Data preprocessing includes cross-platform probe alignment, background correction and normalization. For cross-platform analysis, the primary concern is how to match probes from different platforms. Based on the intrinsic relationships between platforms, we offer three approaches to this issue.

- **Direct match**
  Direct match is used when all probes are identical across microarray platforms.

- **Match by reference IDs**
  Probes from two different platforms can be aligned if they share the same reference ID.

IDs from well-known public databases, for example, UniGene ID or Ensembl ID, can serve as reference ID's, as can any user-defined category.

- **Match by file**
  Users can align probes by providing a probe-mapping file, in which homologous probes are explicitly mapped.

If multiple platforms are involved, normalization within or between arrays of the same platform can be done directly on the raw data before probe alignment. After alignment, the whole data set can be normalized.

Differential analysis

Users can analyze data based on either ratio or intensity. The ratio-based model is $R = \mu + \varepsilon$, where R is the ratio, μ represents the intercept of the ratio of the two groups and ε represents the Gaussian random error. We say two samples are different if μ significantly differs from the null hypothesis.

More than one comparison among groups of data can be requested simultaneously. Furthermore, users may apply "+", "-" and parentheses to make more specific comparisons. For instance, given four groups, "$(group1 + group2) - (group3 + group4)$" computes the global difference between array data supplied in the first two groups compared to array data supplied in the second two groups.

Fold-change analysis, Student's t-test, eBayes-moderated t-test [Smyth, 2004, Smyth et al., 2005], SAM test [Tusher et al., 2001], non-parametric tests (including Wilcoxon rank sum test, Kruskal-Wallis rank sum test and Friedman rank sum test) and ANOVA/ANCOVA are among the choices of algorithms for differential analysis of data in WebArrayDB.

Mixed-effect model ANOVA plays a very important role in microarray data analysis [Churchill, 2002]. ANOVA is capable of dealing with multiple factors. The default model in WebArrayDB is

$$E = \mu + G + P + A + D + S + I + \varepsilon \qquad (1)$$

where E is the observed log-transformed intensity value, μ is the theoretical "real" log-transformed intensity value, ε represents the Gaussian random error with 0 as expected value, and G is the group factor, which leads to effects of interest, e.g. treatment effects. P, A, D, S and I represent effects of *platform*, *array*, *dye*, *sample* and *individual* respectively, among which *array* and *individual* are considered random effect factors. Based on the data to be analyzed, more or fewer factors might be used in specific analysis processes.

Experienced users can define new factors and create complicated analysis models. This enables WebArrayDB to analyze data from virtually any experimental design and thereby to retain relevance as methods continue to evolve.

Other analysis tools

Both raw and differentially analyzed data can be used for further analysis, including hierarchical clustering, correspondence analysis, between group analysis, and plotting using genome position. A variety of high-quality charts in PDF and EPS formats can be produced to visualize analysis results.

Example

Data sources

A demonstration of a cross-platform analysis is used as a training example in every WebArray account. This example uses two publicly available prostate cancer microarray data sets. One set was obtained using a custom made cDNA microarray (20K chip, platform MPMDB ID:42) that contains 19,947 sequence verified PCR-amplified human cDNAs representing 15,495 UniGene clusters [Dhanasekaran et al., 2005] (project MPMDB ID:76). The other was obtained using a commercially available oligonucleotide microarray (Affymetrix U95A array, platform MPMDB ID:9) that contains 12,626 probe sets consisting of 25-base oligonucleotide probes [Welsh et al., 2001] (project MPMDB ID:78). From the two data sets, 49 tumor samples (prostate cancer) and 21 non-tumor samples are analyzed in this example.

Options for analysis

Analysis options selected for this demonstration are illustrated in **Figure 2**. The IDs from the UniGene database (http://www.ncbi.nlm.nih.gov/UniGene) are used to match cDNA clones and Affymetrix probe sets between platforms. Within each study, the median value is used for expression values corresponding to probes of the same UniGene cluster. Genes not mapping to a UniGene cluster present in both microarray platforms are not considered for cross-platform analysis. For integration and normalization of microarray measurements from different platforms, we apply quantile discretization [Warnat et al., 2005]. A common reference sample is used in the two color cDNA microarray study

and the log2 ratios of the intensity values from experimental samples over the common reference sample are calculated for each individual array and used for further analysis. A non-parametric analysis method, the Wilcoxon rank sum test, is used for differential analysis.

Results

A total of 4690 probes are identified as common to both datasets, among which 661 are reported to be differentially expressed between tumor and non-tumor samples at $p < 0.01$, with 267 retained after false discovery rate adjustment by the step-up method of Benjamini-Hochberg (1995). Hierarchical clustering is performed for the top 30 most significant differential expressed gene sets (**Figure 3**). Clustering results show that the samples were separated into two major groups correlating with their biological origin (tumor vs non-tumor instead of their platforms. In general, discriminative gene sets found in two data sets on different platforms are likely to be more reliably characteristic of tumor status than the genes obtained from each individual data set [Warnat et al., 2005].

Implementation

WebArrayDB has been implemented on a LAMP system (a Linux server with Apache, MySQL and Python) in a typical browser/server model (**Figure 4**). In a deployment, the WebArrayDB web server, database server and file server can be located on a single machine or on separate machines. Most modules are written in python (http://www.python.org), while analysis functions are powered by R language (http://www.r-project.org) [R Development Core Team, 2006] and Bioconductor [Gentleman et al., 2004]. Our WebArrayDB is hosted on a Dell server with 4 CPU cores with hyper-threading technology, 24GB of RAM, 1 TB main hard disk and 1 TB hard disk for backup. The configuration will be upgraded depending on the burdens of computation and increases in the data stored.

Parallel computation can be done at two levels:

- Multiple analysis requests from users can be processed simultaneously. In order to avoid too many active requests, WebArrayDB will automatically determine a maximum number of requests that can be processed simultaneously, limiting both the number per user and the total number, while keeping other requests waiting in the queue. The default values can be adjusted by the administrator.
- Even in a single analysis request, computation can be distributed into many processes that run in parallel. The number of processes can be adjusted by the administrator. The package SNOW [Rossini et al., 2003] was adopted for this purpose, so Message Passing Interface (MPI), Parallel Virtual Machine (PVM) or SOCKET can be used for communication in parallel computation.

Although WebArrayDB is presented as a web server on the internet, a package is downloadable for those who want to build their own dedicated servers with Win32 or POSIX (Portable Operating System Interface) on SMP systems or Linux clusters. WebArrayDB is designed as a lightweight database with a user friendly web interface facilitating ease of use for bench scientists. Although a curator is always desirable there is no necessity for one. WebArrayDB is an ideal tool for individual researchers, laboratories, or small research institutes, to store, share and analyze the microarray data. The installation of the WebArrayDB server and maintenance is likely to require only a few hours of assistance of IT staff.

Tutorial and examples

A web-based tutorial, presented in English, Chinese, and Spanish at the WebArrayDB website (http://www.webarraydb.org), shows how to upload data and how to process a simple example. The input data and analysis results used in the tutorial (simple analysis) and this manuscript (complex cross-platform comparison) are available for viewing by all WebArrayDB users. Analysis methods other than the preselected ones can be chosen for these examples, and results of these changes can be viewed and stored in the user-specific accounts. Thus, all new users have the opportunity to familiarize themselves with the powerful capabilities of WebArrayDB by browsing and editing both the simple and the complex examples in the "demo" account upon first entry into the system.

Acknowledgments

This work was made possible by the generous support of Sidney Kimmel, Ira Lechner, Eileen Haag, and Ron Neeley. We also thank Yong Jiang, Krzysztof Studziński, Rocio Canals and Sang-Ho

Choi for testing WebArrayDB, and Fred Long for maintaining the server. This work was performed in the laboratory of Michael McClelland and was supported, in part, by the Prostate Cancer Foundation and Mary Kay Ash Foundation, and National Institutes of Health grants (R01AI034829, R01AI052237, R01CA68822, and U01CA114810), and a grant from the DOD, W81XWH-08-1-0720.

References

Ashburner, M. and Lewis, S., 2002. On ontologies for biologists: the gene ontology–untangling the web. *Novartis Found Symp*, **247**:66–80; discussion 80–3, 84–90, 244–52.

Benjamini, Y. and Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society (Series B)*, **57**:289–300.

Bosotti, R., Locatelli, G., Healy, S., Scacheri, E., Sartori, L., Mercurio, C., Calogero, R., and Isacchi, A., 2007. Cross platform microarray analysis for robust identification of differentially expressed genes. *BMC Bioinformatics*, **8 Suppl 1**:S5.

Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C. A., Causton, H. C., *et al.*, 2001. Minimum information about a microarray experiment (miame)-toward standards for microarray data. *Nat Genet*, **29**(4):365–371.

Bussey, K. J., Kane, D., Sunshine, M., Narasimhan, S., Nishizuka, S., Reinhold, W. C., Zeeberg, B., Ajay, W., and Weinstein, J. N., 2003. Matchminer: a tool for batch navigation among gene and gene product identifiers. *Genome Biol*, **4**(4):R27.

Chen, Q.-R., Song, Y. K., Wei, J. S., Bilke, S., Asgharzadeh, S., Seeger, R. C., and Khan, J., 2008. An integrated cross-platform prognosis study on neuroblastoma patients. *Genomics*, **92**(4):195–203.

Churchill, G. A., 2002. Fundamentals of experimental design for cdna microarrays. *Nat Genet*, **32 Suppl 2**:490–495.

Demeter, J., Beauheim, C., Gollub, J., Hernandez-Boussard, T., Jin, H., Maier, D., Matese, J. C., Nitzberg, M., Wymore, F., Zachariah, Z. K., *et al.*, 2007. The stanford microarray database: implementation of new analysis tools and open source release of software. *Nucleic Acids Res*, **35**(Database issue):D766–D770.

Dhanasekaran, S. M., Dash, A., Yu, J., Maine, I. P., Laxman, B., Tomlins, S. A., Creighton, C. J., Menon, A., Rubin, M. A., and Chinnaiyan, A. M., *et al.*, 2005. Molecular profiling of human prostate tissues: insights into gene expression patterns of prostate development during puberty. *FASEB J*, **19**(2):243–245.

Gautier, L., Cope, L., Bolstad, B. M., and Irizarry, R. A., 2004. affy–analysis of affymetrix genechip data at the probe level. *Bioinformatics*, **20**(3):307–315.

Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., *et al.*, 2004. Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biology*, **5**:R80.

Kapushesky, M., Kemmeren, P., Culhane, A. C., Durinck, S., Ihmels, J., Körner, C., Kull, M., Torrente, A., Sarkans, U., Vilo, J., *et al.*, 2004. Expression profiler: next generation–an online platform for analysis of microarray data. *Nucleic Acids Res*, **32**(Web Server issue):W465–W470.

Killion, P. J., Sherlock, G., and Iyer, V. R., 2003. The longhorn array database (lad): an open-source, miame compliant implementation of the stanford microarray database (smd). *BMC Bioinformatics*, **4**:32.

Larkin, J. E., Frank, B. C., Gavras, H., Sultana, R., and Quackenbush, J., 2005. Independence and reproducibility across microarray platforms. *Nat Methods*, **2**(5):337–344.

Moreau, Y., Aerts, S., Moor, B. D., Strooper, B. D., and Dabrowski, M., 2003. Comparison and meta-analysis of microarray data: from the bench to the computer desk. *Trends Genet*, **19**(10):570–577.

Pan, F., Kamath, K., Zhang, K., Pulapura, S., Achar, A., Nunez-Iglesias, J., Huang, Y., Yan, X., Han, J., Hu, H., *et al.*, 2006. Integrative array analyzer: a software package for analysis of cross-platform and cross-species microarray data. *Bioinformatics*, **22**(13):1665–1667.

Parkinson, H., Kapushesky, M., Shojatalab, M., Abeygunawardena, N., Coulson, R., Farne, A., Holloway, E., Kolesnykov, N., Lilja, P., Lukk, M., *et al.*, 2007. Arrayexpress–a public database of microarray experiments and gene expression profiles. *Nucleic Acids Res*, **35**(Database issue):D747–D750.

R Development Core Team, 2006. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Rossini, A., Tierney, L., and Li, N., 2003. Simple parallel statistical computing in R. In *UW Biostatistics working paper series, WA Paper 193*. University of Washington.

Saal, L. H., Troein, C., Vallon-Christersson, J., Gruvberger, S., Borg, A., and Peterson, C., 2002. Bioarray software environment (base): a platform for comprehensive management and analysis of microarray data. *Genome Biol*, **3**(8):SOFTWARE0003.

Shippy, R., Sendera, T. J., Lockner, R., Palaniappan, C., Kaysser-Kranich, T., Watts, G., and Alsobrook, J., 2004. Performance evaluation of commercial short-oligonucleotide microarrays and the impact of noise in making cross-platform correlations. *BMC Genomics*, **5**(1):61.

Smyth, G. K., 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology*, **3**:Iss. 1, Article 3.

Smyth, G. K., 2005. Limma: linear models for microarray data. In Gentleman, R., Carey, V., Dudoit, S., Irizarry, R., and Huber, W., editors, *Bioinformatics and Computational Biology Solutions using R and Bioconductor*, pages 397–420. Springer. New York.

Smyth, G. K., Michaud, J., and Scott, H., 2005. The use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics*, **21**(9):2067–2075.

Stoyanova, R., Querec, T. D., Brown, T. R., and Patriotis, C., 2004. Normalization of single-channel dna array data by principal component analysis. *Bioinformatics*, **20**:1772–1784.

Tárraga, J., Medina, I., Carbonell, J., Huerta-Cepas, J., Minguez, P., Alloza, E., Al-Shahrour, F., Vegas-Azcrate, S., Goetz, S., Escobar, P., *et al.*, 2008. Gepas, a web-based tool for microarray data analysis and interpretation. *Nucleic Acids Res*, **36**(Web Server issue):W308–W314.

Troein, C., Vallon-Christersson, J., and Saal, L. H., 2006. An introduction to bioarray software environment. *Methods Enzymol*, **411**:99–119.

Tusher, V. G., Tibshirani, R., and Chu, G., 2001. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A*, **98**(9):5116–5121.

Warnat, P., Eils, R., and Brors, B., 2005. Cross-platform analysis of cancer microarray data improves gene expression based classification of phenotypes. *BMC Bioinformatics*, **6**:265.

Welsh, J. B., Sapinoso, L. M., Su, A. I., Kern, S. G., Wang-Rodriguez, J., Moskaluk, C. A., Frierson, H. F., and Hampton, G. M., 2001. Analysis of gene expression identifies candidate markers and pharmacological targets in prostate cancer. *Cancer Res*, **61**(16):5974–5978.

Xia, X., McClelland, M., and Wang, Y., 2005. Webarray: an online platform for microarray data analysis. *BMC Bioinformatics*, **6**:306.

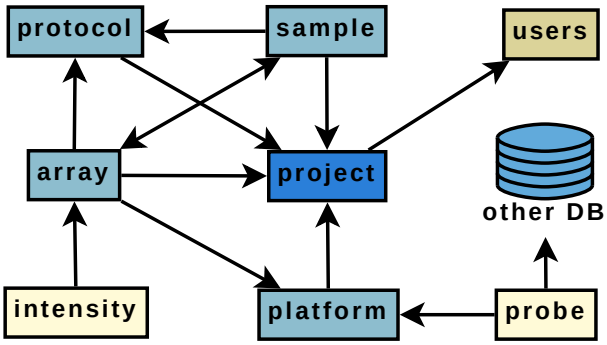


Figure 1: Information organization in WebArrayDB.

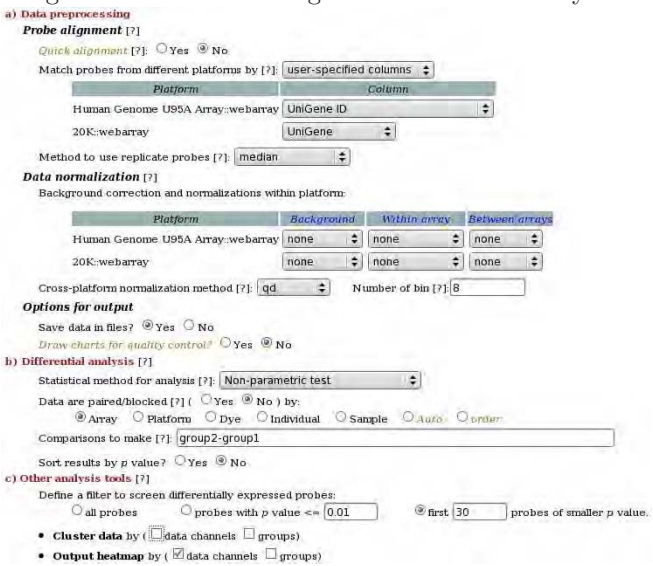


Figure 2: Options selected in an analysis of two publicly available prostate cancer microarray data sets. See text for details.

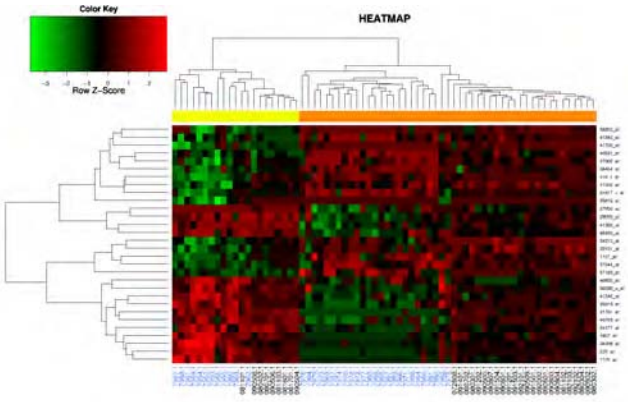


Figure 3: Heat map of the 30 most significantly differentially expressed probes between tumor and non-tumor samples.
The tumor samples are marked at the top of the plot by a brown bar and the non-tumor group by a yellow bar. Arrays of the 20K platform are named in blue font at the bottom of the plot, Affymetrix U95A arrays in black font.

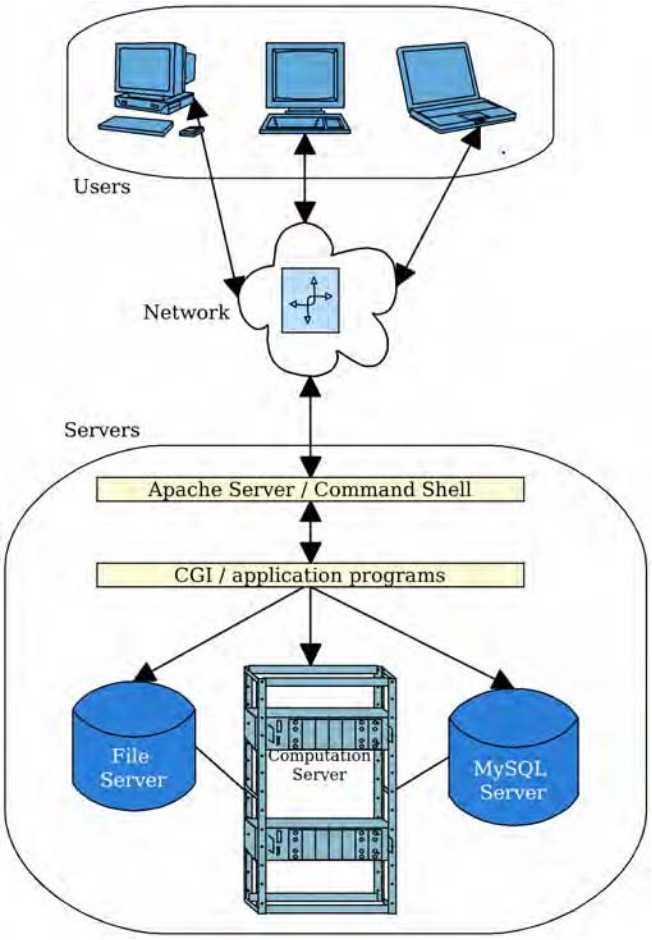


Figure 4: Architecture of WebArrayDB.

Evaluating Oligonucleotide Properties for DNA Microarray Probe Design

Xiao-Qin Xia [a,e,*], Zhenyu Jia [b,e], Steffen Porwollik [a], Fred Long [a], Claudia Hömme [c], Kai Ye [d], Carsten Müller-Tidow [c], Michael McClelland [a,*], and Yipeng Wang [a,b,*],

[a] *Vaccine Research Institute of San Diego, 10835 Road to the Cure, Suite 150, San Diego, CA 92121, USA*

[b] *Department of Pathology & Laboratory Medicine, University of California, Irvine, CA 92697, USA*

[c] *Department of Medicine A, Hematology and Oncology, University of Münster, Domagkstr. 3, 48129 Münster, Germany*

[d] *Molecular Epidemiology, Medical Statistics and Bioinformatics, Leiden University Medical Center, The Netherlands*

[e] *These authors contributed equally.*

Abstract

Most current microarray oligonucleotide probe design strategies are based on probe design factors (PDFs), which include probe hybridization free energy (PHFE), probe minimum folding energy (PMFE), dimer score, hairpin score, homology score, and complexity score. The impact of these PDFs on probe performance was evaluated using four sets of microarray comparative genome hybridization (aCGH) data, which included two array manufacturing methods and the genomes of two species. Since most of the hybridizing DNA is equimolar in CGH data, such data are ideal for testing the generally hybridization properties of almost all candidate oligonucleotides. In all our datasets, PDFs related to probe secondary structure (PMFE, hairpin score and dimer score) are the most significant factors linearly correlated with probe hybridization intensities. PHFE, homology and complexity score are correlating significantly with probe specificities, but in a non-linear fashion. We developed a new probe design factor, pseudo probe binding energy (PPBE), by iteratively fitting di-nucleotide po-

sitional weights and di-nucleotide stacking energies until the average residue sum of squares (ARSS) for the model was minimized. PPBE showed a better correlation with probe sensitivity and a better specificity than all other PDFs, although training data are required to construct PPBE model first prior to designing new oligonucleotide probes. The physical properties that are measured by PPBE are as yet unknown but include a platform-dependent component. A practical way to use these PDFs for probe design is to set cut-off thresholds to filter out bad quality probes. Programs and correlation parameters from this study are freely available to facilitate the design of DNA microarray oligonucleotide probes.

Key words: microarray, probe design, oligonucleotide

Introduction

Microarray technology surveys many thousands of genes to investigate gene expression [1], transcription factor binding profiles [2–5], DNA methylation profiles [4,6], comparisons of DNA copy number [5] and comparative genomic sequencing [7].

Oligonucleotide probes provide higher hybridization specificity than longer PCR products [8–10]. Falling costs of oligonucleotide synthesis, along with the development of new microarray manufacture technologies, such as the NimbleGen maskless array synthesizer [11] and Agilent's ink-jet oligonucleotide synthesizer make custom long (¿ 50 bases) oligonucleotide arrays possible for many experimental applications. Optimal probe design algorithms are consequently desirable.

Hybridization on an array can be explained by several interconnected processes, including the affinity of a target for a probe, the formation of stem-loop structures of a probe, the formation of secondary structures (loops and helices) of a target, and probe-to-probe dimerization [12–16]. There are a variety of factors governing these processes, including probe hybridization energy

* Corresponding authors.
Email addresses: xqxia70@gmail.com (Xiao-Qin Xia), mcclelland.michael@gmail.com (Michael McClelland), yipengw@gmail.com (Yipeng Wang).

(PHFE) [17], probe minimum folding energy (PMFE) [18], probe dimer and hairpin scores [19], as well as homology and complexity scores [20]. Most of the current oligonucleotide probe design software packages estimate these properties [20–28].

To systematically and quantitatively study how these factors influence probe performance in microarrays, we collected a large amount of array CGH microarray data and used these data to evaluate the utility of each probe design factor (PDF) for probe selection. Using aCGH data, a novel probe design factor, pseudo probe binding energy (PPBE), was developed. PPBE is more accurate in predicting probe performance than all other factors and can thus be used for iterative improvement of the choice of oligonucleotides on the array. While the specific physical properties measured by PPBE remain unknown, they encompass platform-specific parameters.

Methods

Microarray CGH Data Sets

Four comparative genome hybridization microarray data sets were used in the study (**Table 1**). Human genomic DNA (data sets 1, 2 and 4) and *Salmonella* genomic DNA (data set 3) samples were hybridized to their corresponding arrays. The array platforms include NimbleGen arrays (3' end of oligos is linked to the solid phase) and in-house spotted oligonucleotide arrays (5' end of oligos is linked to the solid phase). The majority of probes on the arrays we use are 50 nucleotides in length. However, there are also probes of different length, e.g., there are 9989 of 46-mer probes and 4721 of 55-mer probes on the array for data set 4. We found that the correlations of PDFs to probe sensitivities for these probes are very similar to those of the 50-mer probes (data not shown). In order to make data comparable across platforms, only data from 50-mer oligonucleotide probes were used. Hybridization intensity values were natural log transformed before fitting the linear models.

Samples that were hybridized to the arrays included human and *Salmonella* genomic DNA. Data

set 3 used pooled *Salmonella* genomic DNA *Xba*I restriction fragments, representing half of the genome in three-fold excess, in one channel, and whole genomic DNA in the other. Data set 4 contains 205 replicates of human lung tissue genomic DNA hybridizations which were used as control channel in two-color hybridizations experiments.

Probe Design Factors

The following DNA microarray probe design factors were included in this study.

Probe hybridization free energy (PHFE)

PHFE was calculated based on the di-nucleotide stacking energies.

$$PHFE = \varepsilon_{head} + \sum_{k=1}^{n-1} \varepsilon(b_k, b_{k+1}) + \varepsilon_{tail}$$

where n is the oligonucleotide length, $\varepsilon(b_k, b_{k+1})$ is the kth position di-nucleotide stacking energy, and ε_{head} and ε_{tail} are the terminal nucleotide stacking energies. The salt concentrations for the calculations were set to 1M Na+, 0M Mg++, and the temperature was set to 40, 50 or 60℃ for the computation of PHFE. The di-nucleotide stacking energies are computed according to SantaLucia [17] and shown in **Supplementary Table 1**.

Pseudo Probe Binding Energy (PPBE)

For a probe sequence $(b_1, b_2, \ldots, b_n)$ with n bases, the PPBE model is parameterized by di-nucleotide stacking energies ε and position dependent weights ω, $PPBE = \varepsilon_{head} + \sum_{k=1}^{n} \omega_k \varepsilon(b_k, b_{k+1}) + \varepsilon_{tail}$. The position-dependent weight ω is first estimated by fitting the linear model, employing di-nucleotide stacking energies (as used in the PHFE model) as initial values. Then, with the same linear model fitting scheme, the pseudo di-nucleotide stacking energies ε are approximated by treating previously estimated weights as known quantities. Such process of "reciprocal" estima-

tion was iteratively carried out three times, at which point the ARSS for the PPBE model reached its minimum or near-minimum (see also the Linear Modeling section below, and **Figure 1A**).

Probe minimum folding energy (PMFE)

PMFE is the minimum folding energy of a single strand DNA sequence and represents the stability of the secondary structure of a given sequence. PMFE were computed by using the MFOLD program [18]. The program *hybrid-ss-min* was downloaded from
http://www.bioinfo.rpi.edu/applications/hybrid/download.php
and executed on GNU/Linux. The parameters were set as DNA-DNA hybridization, 1M Na+, 0M Mg++, and the temperature was set to 40, 50 or 60℃ for calculation of PMFE.

Probe dimer score, hairpin score

The calculation of the probe dimer score and the hairpin score was described as part of the AutoDimer program based on a sliding algorithm [19]. For screening probe dimers, two probe sequences are incrementally overlapped, and the presence or absence of base pairing is evaluated and tabulated. A dimer score value was then determined by combining the number of Watson-Crick base pairs (+1) with mismatches (-1).

Hairpin secondary structures were screened by using the probe sequence to check for the presence of 4 and 5 base loops. A minimum of a 2-base stem were deemed to be necessary in a hairpin structure. Hairpin scores were sums of matched base pairs (+1) in hairpin stems where mismatches are not permitted.

Homology Score

The homology score for each oligonucleotide estimates the degree of cross hybridization, and is based on a BLAST search of the input sequence against a species-specific database. The calculation of the homology score was similar to the one used in the OligoWiz program [20].

$$Homology\,Score \;=\; \frac{100 \times L - \sum_{i=1}^{L} \max\left(B_{1i}, \ldots, B_{mi}\right)}{100 \times L}$$

where L is the length of the oligonucleotide, m is the number of Blast hits considered in position i of the oligonucleotide and $B = \{B_{1i}, \cdots, B_{mi}\}$ is the bit score in position i.

Oligonucleotides with 100% identity to any considered BLAST hit along the full length gets a score of 0. A score value will be assigned to oligonucleotides that have no perfect homology to any considered BLAST hit. Percentages of identity lower than 70% or shorter than 15bp were removed, resulting in perfect homology scores of 1.

Complexity Score

Complexity scores were calculated for estimating the degree of common sequence fragments in a given oligonucleotide, as described in the OligoWiz program [20]. The information content can be calculated by the following equation:

$$I\left(w\right) \;=\; \frac{n\left(w\right)}{nt}\left(\log_2 \frac{n\left(w\right) \times 4^{l(w)}}{nt}\right)$$

where $n\left(w\right)$ is the number of occurrences of a pattern in the genome, $l\left(w\right)$ the pattern length, *nt* is the total number of patterns found in DNA sequences present in the target pool, for example, the whole genome in an array comparative genomic hybridization. The following equation was used

to calculate the complexity score for each oligonucleotide probe:

$$Complexity\ Score\ =\ 1-norm\left(\sum_{L-l(w)+1}^{i=1} I\left(w_i\right)\right)$$

where L is the length of the oligonucleotide, w_i is the pattern in position i and norm is a function that normalizes the summed information to a value between 1 and 0 by dividing them by the maximum value. A complexity score of 0 indicates an oligonucleotide with very low complexity. Pattern lengths of 2, 5, 8 and 11 bases were tested in this study.

Oligonucleotide Specificity and Reproducibility

Data set 3, with known expected oligonucleotide signal ratios (three fold changes) between the two channels, was used for estimating oligonucleotide probe specificity. The observed ratios were *log* base 2 transformed for further analysis. Coefficient of variation (cv) was used for estimating probe reproducibility.

Linear Modeling and Model Validation

R language (http://www.r-project.org) was used for linear modeling [29–31]. In the four microarray data sets, simple linear models were used to evaluate each individual probe design factor and multivariate models were used to estimate all probe design factors together.

The Average Residue Sum of Squares (ARSS), which reflects the model fitness, was defined as $r = \frac{\sum_{i=1}^{n}\left(g_i-g_i^*\right)^2}{n}$, where g_i was the observed *ln*-transformed intensity for probe i, g_i^* was the predicted *ln*-transformed intensity for probe i, and n was the number of probes. For model selection, the stepAIC function in the MASS package (http://www.r-project.org) was used to reduce the full model to the optimal one. This Akaike information criterion (AIC) is a measure of the quality of fit of an estimated statistical model and balances the complexity of an estimated model with the accuracy with which the model fits the data [32].

The models were validated in two ways: within one data set and across different data sets. In both cases, the leave-many-out cross-validation [33] was used. Within-dataset validation uses half of the data from one data set to train the models and the other half for testing of the models. Cross-dataset validation uses different data sets, which may vary in array platforms and sample species, for training and testing.

Results

Microarray CGH Data Sets

Array CGH data is a valuable source for studying microarray oligonucleotide probe performance because it can be assumed that most of the probes in these experiments hybridize to approximately equimolar target amounts, resulting in relatively uniform hybridization signals. Four large aCGH data sets on different array platforms, with a total of 657,646 of 50-mer oligos and 219 samples, were used in this study to evaluate probe design factors and to develop new algorithms (see **Table 1**).

Correlation of Individual Probe Design Factors (PDFs) with Probe Hybridization Intensities

The models examined are all presented in the methods section and will not be repeated here. All ten probe design factors (PDFs), i.e., probe hybridization free energy (PHFE), probe minimum folding energy (PMFE), hairpin score, probe dimer score, homology score, complexity score (2 bases), complexity score (5 bases), complexity score (8 bases), complexity score (11 bases), and pseudo probe binding energy (PPBE), showed highly significant correlation with probe hybridization intensities, as shown in **Figure 2** (data set 1) and **Supplementary Figure 1** (data set 2, 3 and 4). The correlation coefficients (r), ARSS, intercepts and slopes for these linear regression models are listed in **Table 2** and **Supplementary Table 2**.

The average residue sum of squares (ARSS) values of linear models based on individual PDFs were

compared, as shown in **Figure 3**. Among these factors, PPBE generated the lowest ARSS, suggesting that this factor is superior to the traditional factors in correlating with probe hybridization intensity. PPBE was modeled by iteratively fitting di-nucleotide stacking energies and positional weights, with the conventional di-nucleotide stacking energies as initial values. The ARSS values from the PPBE model tend to stabilize after three cycles of iterative fitting of each of positional weights and pseudo di-nucleotide stacking energies (**Figure 1** and **Supplementary Figure 2**). The positional weights and pseudo di-nucleotide stacking energies generated from the different data sets are entirely different, reflecting the empirical nature of the model. The positional weights and pseudo stacking energies for PPBE models from different data sets are listed in **Supplementary Table 3 and 4**, the positional weights illustrates the effect of the distance of the dinucleotide to the solid phase. The positional weights of data set 2 and data set 4, for example, showed inverse correlation for the distance to the probe's 5' end, which may due to the fact that these platforms differed in the ends of oligos that were linked to the solid phase (5' versus 3').

The best individual traditional factors are PMFE, dimer score and hairpin score in most data sets. All these three PDFs showed that less stable probe secondary structure positively correlates with probe hybridization intensity, suggesting that the formation of secondary structure can severely hinder the probe hybridization capabilities.

PHFE's linear correlation with probe hybridization intensity was less significant, suggesting that hybridization behavior on microarrays might be different from that in solution. Moreover, quadratic rather than linear relationships were observed for data set 1 and 3 and the mode (the peak points shown in **Figure 2A** and **Supplementary Figure 1-2A**) varies among these two data sets, suggesting that hybridization conditions were not the same for the two data sets. We tried to use quadratic equations to fit the data set 1 and 3, but the ARSS values generated from these models were bigger than those obtained using simple linear models (data not shown). This is probably due to the fact that the majority of PHFE data points are clustered within a very narrow range, where the relationship between PHFE and intensities may be better described by a linear equation. In future studies

once there are sufficiently large data sets with a higher PHFE data spread across a wider range of values, more advanced models can be applied to scrutinize the relationship between PHFE and hybridization intensities in a non-linear fashion.

Blast score and complexity scores (2, 5, 8, 11 bases) correlated least significantly with the probe hybridization intensity among the PDFs tested. No obvious differences were observed among the scores obtained for 2, 5, 8 and 11 bases when correlating them with probe hybridization intensity (**Table 2**).

Among all four data sets, PPBE, PMFE, dimer score, and hairpin score showed positive correlation with probe hybridization intensity, and are therefore the more reliable indicators of probe sensitivity. The other PDFs displayed inconsistencies in correlation for different data sets. For example, PHFE is positively correlated with probe intensity in data sets 2 and 3, but is negatively correlated with probe intensity in data sets 1 and 4. More complex models might be developed for blast score and complexity scores (2, 5, 8, 11 bases), but that is beyond the scope of this paper.

As shown in **Supplementary Table 2**, enormous variations were observed among individual data sets for the trend coefficients (e.g., intercept and slope), possibly due to differences in array manufacture, sample and array processing, and other factors.

The values of PHFE and PMFE are dependent on parameters such as hybridization temperature and concentrations of sodium, most of which were unavailable to us. However we computed PHFE and PMFE using various potential parameters, and changes in parameters did not cause significant differences in correlation assessments, the average difference of ARSS value are 0.0058 (0.010 for PHFE and 0.001 for PMFE) among different temperature setting. 60℃ was used for the PHFE computation presented and 40℃ was used for PMFE computation presented for all data sets because they slightly outperformed other temperatures.

Multi-variate Linear Modeling

For each data set, a multi-variate linear model with PPBE (W. PPBE model) was built based on all PDFs for predicting probe hybridization intensity and comparing the significance of the individual PDFs. This multi-variate model showed significant improvement over all individual models based on each individual PDF (note the significantly diminished ARSS values in **Figure 3** and **Supplementary Figure 3**). The W. PPBE model parameters are in **Supplementary Table 5**.

Increasing the number of free parameters obviously improves the fit. On the other hand, overfitting is very likely to happen and reduces or destroys the ability of the model to generalize beyond the data it is built upon. The Akaike information criterion (AIC) is an operational way of trading off the complexity of an estimated model against how well the model fits the data [32]. It not only rewards improvement of fit, but also includes a penalty that is an increasing function of the number of estimated parameters and thereby discourages over-fitting. In this study, stepwise selection with AIC was used to search for the optimal model which only contains covariates (individual PDFs) related to the outcome (probe hybridization intensity). Stepwise model selection analysis showed that all PDFs contributed to the prediction of probe hybridization intensity in all data sets with only one exception in which the complexity score (2 bases) was not significant in data set 1 (**Supplementary Figure 4**). The most significant factor is PPBE, followed by PMFE in all data sets. The order of significance of other PDFs may vary among different data sets.

Generality of Linear Models

Two multi-variate models, the W. PPBE model (includes all factors) and the W/O PPBE model (including all factors except PPBE), were developed using a training data set and tested on independent data sets to determine if the models can be reliably used as a probe design tool.

Applying within-dataset validation, **Figure 4** illustrates that the models developed from the training set can predict the performance of oligos in the test set almost as accurately as it can predict

performance in the training set. W. PPBE model outperformed W/O PPBE in all cases suggesting that PPBE is a reliable factor although it is generated by an empirical approach.

Cross-dataset validations (**Supplementary Table 6**) resulted in extremely high ARSS values in the test data sets when the W/O PPBE and W. PPBE models were applied, even if the array manufacture technique and sample species were identical between test and training set. The complex multi-variate models developed from one data set can therefore not be directly and simply applied on other data sets. The adverse performance was not caused by PPBE, as there were no obvious differences between W/O PPBE and W. PPBE models. The substantial variations in correlation intercepts and slopes for each individual PDF, as observed in **Supplementary Table 2**, severely hinder the cross-dataset probe intensity predictions using multi-variate linear models.

Probe Specificity

Probe specificity is a measurement of the capability of a probe to discriminate between its specific target sequences in the context of a complex set of non-specific sequences. In a two-channel hybridization experiment, if one channel includes the target sequence and the other does not, then the probe with specificity for the target can be expected to yield a high ratio of hybridization signal intensity between the two channels, which is a measure of probe specificity in the mixture.

We estimated the oligonucleotide specificity using Data Set 3, where the targets in one channel included a three-fold over-representation of approximately half of the *Salmonella* genome and three-fold under-representation for the other half of the genome. Therefore there are three fold differences in the target concentration between the two channels for all probes and the expected hybridization ratio is 3 for specific hybridization. This was achieved by *Xba*I-digestion of stationary phase *Salmonella enterica* sv Typhimurium LT2 genomic DNA, separation of the seven fragments using pulsed field gel electrophoresis, capturing those fragments and pooling the six smaller fragments, while keeping the big fragment separate. Genomic DNA preparations from sta-

tionary phase LT2 were then supplemented either with the big fragment, or with the pooled six smaller fragments, creating overrepresentations of the different halves of the genome.

Probes with stronger hybridization intensities displayed better specificity (**Figure 5A**). When each individual PDF and the predicted probe hybridization intensities were compared with the observed ratios, significant correlation was detected between the ratios and all the factors (**Supplementary Figure 6**), most significantly for PHFE, PMFE, PPBE and Complexity Score (8 bases). The Pearson correlation coefficients are listed in **Supplementary Table 7**. It is interesting to note that PHFE is significantly and positively correlated with probe specificity. Probes with low PHFE values displayed both low specificity and relatively low sensitivity (as shown in **Supplementary Figure 1-2**).

As shown in **Supplementary Figure 5**, the relationships between *log*2 based ratios and some PDFs seem to be non-linear. For the sake of simplicity, only linear equations were considered in the current study.

Probe Reproducibility

Data set 4, which includes 205 replicated hybridizations, was used to estimate probe reproducibility using coefficient of variation (cv). High probe reproducibility (corresponding to low cv values) is positively correlated with the observed probe hybridization intensities (**Figure 5B**). When examined individually, each PDF shows a significant but distinct level of associations with cv (**Supplementary Figure 6**). PPBE and PHFE are the most significant factors. Correlation coefficients are listed in **Supplementary Table 7**. Note that only linear equations were considered for this reproducibility survey.

Software

Programs for computing of PHFE, PMFE, probe dimer score and hairpin score, blast score and complexity score were written in Python. All programs, including parameters for computation, are

freely available upon request.

Discussion

Microarray probe hybridization signals are determined by the equilibrium of probe-target complex formation and probe-probe hybridization capability, and are also influenced by non-specific binding from the complex target. The probe design factors (PDFs) we studied here covered these three aspects.

Although Affymetrix Chips are designed for one-sample-for-one-array, it is very common to apply multiple samples on a single array from customized platforms, including in-house spotted arrays and many Nimblegen arrays and we took advantage of this fact. The natural log transformed intensity values from multiple arrays were averaged for each probe to minimize variation caused by sample processing and hybridization. One advantage of our datasets for comparing probe performance is that genomic DNA samples have targets at the same or similar concentrations, allowing a comparison of probe performance under similar target concentrations.

Linear models were selected to model the relationships between individual PDFs and probe performance based on our observation that most scatter plots generated from multiple data sets consistently showed a linear relationship. The actual relationships may be far more complex, nevertheless, for a practical point of view, linear models are easy to handle and generate more accurate predictions based on model diagnosis with ARSS than more complex models [34]. The finding of these correlations is a useful first step in trying to understand the physical phenomena, which are clearly not subsumed in all the parameters currently in use. In future research, we plan to identify more advanced models (for example non-linear association models) which may reduce the ARSS we have achieved in the current study.

Probe minimum folding energy (PMFE), dimer score and hairpin score were the factors used to estimate the probe-probe hybridization capability. Of all the traditional PDFs (all factors except

PPBE), PMFE correlates most significantly with probe hybridization intensity in all four data sets, followed by dimer score and hairpin score in most data sets. Although these three PDFs contain redundant information for estimation of the probe-probe hybridization capabilities, they can not be simply replaced by each other as shown in the stepAIC analysis, which optimizes the complexity of the model versus the fit [32]. All three PDFs therefore deliver unique information that needs to be considered for probe design.

Probe hybridization free energy (PHFE) is a long-established parameter for measuring probe-target hybridization capability in solution. In our study, PHFE was not as reliable in predicting probe hybridization intensity as other factors (PMFE, dimer score and hairpin score), which may be largely due to the linkage of probes to a solid phase in microarray hybridization. To compensate for the effect of one end of the probe being attached to the matrix, we introduced PPBE which modifies the PHFE calculation by adding a positional weight parameter and iteratively fitting positional weights and di-nucleotide stacking energies. PPBE showed much better capabilities of predicting probe hybridization than all other PDFs and tremendous improvement over PHFE. The drawback of PPBE is that it is platform-dependent and preliminary aCGH data is required for developing the PPBE model prior to application. The quality of the training data is critical for the construction of an accurate PPBE model. There are many factors that may result in bad quality arrays, such as bad sample quality, bad hybridization, etc. To solve these problems, we suggest that CGH be performed using normal genomes without copy number variation, and multiple hybridizations with each of the dyes to be used would be desirable to minimize the noise caused by sample processing.

Both PMFE and PHFE are sodium-dependent. Generally, changes in free energy are linearly correlated to log-transformed sodium concentration [17], which has been confirmed by us on the Mfold web server [18] for PMFE and PHFE. That means all the oligonucleotide PMFE/PHFE values will change in the same proportion if the sodium concentration changes. Subsequently, these changes will be cancelled out by adjusting of related coefficients in linear models. Therefore, changes in sodium concentration had no influence on the significance of linear modeling.

The PPBE model is empirical by nature, similar to the positional-dependent-nearest-neighbor (PDNN) model which was designed for the Affymetrix array platform [34], whose parameters similarly need to be empirically estimated based on hybridization data and significantly vary among different Affymetrix array platforms. At this stage, we do not understand the physical properties governing the parameters, but present a practical approach to optimizing oligo design.

The position-dependence of the weighting factors is a conspicuous feature in such models. In previous work, the sensitivity profiles of base C and base A change in a parabola-like fashion along the 25-base probe sequence, while the same terms for G and T change monotonically [35–38]. The overall position weighting factors changes roughly as the curvature of a parabola with peak and shape varying across different GeneChip platforms [14,34,39]. Our data reveal weight distribution patterns different from this previous work. Our data were obtained on two types of platforms: Nimblegen *in situ* synthesized oligonucleotide arrays and a spotted oligonucleotide array. For three Nimblegen platforms, the weights change linearly for the first 35$\sim$45 bases or so from the 3' end and get weaker at the free end (**Figure 1B**, **Supplementary Figure 2B** & **2E**). In contrast, a parabola-like curve is observed on the other platform (**Supplementary Figure 2H**). Although it is not the object of this article to explore a physical explanation for these differences, we point out some facts that may be important in further studies:

- We are using platforms of 50-mer probes, while the quoted previous work used 25-mer Affymetrix GeneChip platforms. Lengthening of the sequence on the platform inevitably reduces the importance of each single base or position, and weakens the position-dependence.
- Unlike Affymetrix platforms and Nimblegen platforms, the probes of the spotted array in this study are linked to the array at the 5 end, and there are no terminal oligonucleotide linkers between probes and the array surface. This impact of this difference is unknown, but it may reduce the freedom of a probe and even its effective length, leading to a pattern of position-dependence similar to platforms of less probe length, e.g. Affymetrix platforms.

For the fitting of the PPBE model, it is not critical whether weights or energies were fitted first. Either way, the final converged models reach similar ARSS values, the average difference is less than 0.005 in ARSS value. The final weights and pseudo stacking energies are similar as well. We began to fit the models with the conventional di-nucleotide stacking energies simply because the modes reached convergence faster. The di-nucleotide stacking energies may express a relevant part of the physical properties underlying the model. It is possible that the di-nucleotide stacking energies may express a relevant part of the physical properties underlying the model; however, further evidence is required to confirm this speculation.

Blast and complexity scores reflect occurrences of sequence segments similar to the probe, and are used for evaluating probe specificity. It would be simpler and easier to use cut-off thresholds for these PDFs to filter out bad quality probes. In this study we applied four different patterns for the complexity score calculation, which are based on 2, 5, 8 and 11 base patterns. The complexity score (8 bases) showed better correlation with probe specificity than other complexity score patterns and blast score.

Langmuir isotherm oriented models were not included in our studies. Although Langmuir model was initially developed for adsorption of gases on glass surfaces [40], its variations have been widely applied in researches on hybridization of oligonucleotides on DNA microarrays [13–16,41]. In these models, the hybridization signal intensities were in essence divided into two parts: the hybridization of the probe with its perfect-matching target and the background noise. Although such models fit hybridization intensity values well for spike-in genes and corresponding targets with controlled concentrations, they are of less help in screening probes for microarray design because these models for microarray design are based on the equilibrium constant, or equivalently, the change of standard Gibbs free energy ΔG°, which is a PDF of less sensitivity and specificity in comparison to PMFE and PPBE in our study. In contrast, platform-dependent empirical models based on pseudo free energies and position weights can make predictions very close to the observed hybridization intensities [34,39]. This fact encouraged us to explore pure empirical models

in microarray design.

In summary, we used aCGH as a model system to study the correlation between individual PDFs and probe performance during microarray hybridization. These individual correlations can be used as guidance for designing microarray probes for other complex experimental setups such as gene expression analysis. In gene expression microarray hybridization, non-specific binding, probe-targets complex formation and probe-probe binding capability will all be influenced by the varying concentrations of the targets. Systematic study of probe performance in such systems is beyond the scope of this study.

If preliminary aCGH data is available, a complex multi-variate linear model including factor PPBE can be developed and used for refining arrays. The model can predict a probe hybridization intensity value which will be an indicator of probe quality. Higher predicted intensity values will be equivalent to higher sensitivity, improved specificity and reproducibility. In practice, this strategy can be used for improving an existing array platform by replacing bad probes or by expanding the array by selecting probes predicted to perform well.

If aCGH data are unavailable for microarray platform design, we suggest using each individual PDF to filter or rank probes instead of using a complex model, because the coefficient parameters (intercept and slopes) vary significantly among different data sets/platforms. PMFE, hairpin score and probe dimer score can be used to rank probe qualities. PHFE, blast score and complexity score can be used to filter probes with low specificity. We have provided all correlation parameters generated from four data sets to be used as a guideline for filtering or ranking probes. All the programs for calculating individual PDFs are also available from the authors.

ACKNOWLEDGEMENTS

We thank the following for their support to MM: The Prostate Cancer Foundation, the Mary Kay Ash Foundation, NIH grants R01CA68822, U01CA114810, R01 AI052237, R01 AI073971,

1R01AI075093 and a grant from the DOD Breast Cancer Research Program, BC073899.

Conflict of interest statement. None declared.

Supplementary Data

Supplementary data are available at NAR Online.

References

[1] Ramsay, G. (Jan, 1998) Dna chips: state-of-the art.. *Nat Biotechnol*, **16**(1), 40–44.

[2] Hayakawa, J., Mittal, S., Wang, Y., Korkmaz, K. S., Adamson, E., English, C., Ohmichi, M., Omichi, M., McClelland, M., and Mercola, D. (Nov, 2004) Identification of promoters bound by c-jun/atf2 during rapid large-scale gene activation following genotoxic stress.. *Mol Cell*, **16**(4), 521–535.

[3] Hoemme, C., Peerzada, A., Behre, G., Wang, Y., McClelland, M., Nieselt, K., Zschunke, M., Disselhoff, C., Agrawal, S., Isken, F., Tidow, N., Berdel, W. E., Serve, H., and Mller-Tidow, C. (Mar, 2008) Chromatin modifications induced by pml-raralpha repress critical targets in leukemogenesis as analyzed by chip-chip.. *Blood*, **111**(5), 2887–2895.

[4] Wang, Y., Hayakawa, J., Long, F., Yu, Q., Cho, A. H., Rondeau, G., Welsh, J., Mittal, S., Belle, I. D., Adamson, E., McClelland, M., and Mercola, D. (Nov, 2005) ”promoter array” studies identify cohorts of genes directly regulated by methylation, copy number change, or transcription factor binding in human cancer cells.. *Ann N Y Acad Sci*, **1058**, 162–185.

[5] Peeters, J. K. and derSpek, P. J. V. (2005) Growing applications and advancements in microarray technology and analysis tools.. *Cell Biochem Biophys*, **43**(1), 149–166.

[6] Wang, Y., Yu, Q., Cho, A. H., Rondeau, G., Welsh, J., Adamson, E., Mercola, D., and McClelland, M. (Aug, 2005) Survey of differentially methylated promoters in prostate cancer cell lines.. *Neoplasia*, **7**(8), 748–760.

[7] Herring, C. D. and Palsson, B. . (2007) An evaluation of comparative genome sequencing (cgs) by comparing two previously-sequenced bacterial genomes.. *BMC Genomics*, **8**, 274.

[8] Relgio, A., Schwager, C., Richter, A., Ansorge, W., and Valcrcel, J. (2002) Optimization of oligonucleotide-based dna microarrays.. *Nucleic Acids Res*, **30**(11), e51.

[9] Chou, C.-C., Chen, C.-H., Lee, T.-T., and Peck, K. (2004) Optimization of probe length and the number of probes per gene for optimal microarray analysis of gene expression.. *Nucleic Acids Res*, **32**(12), e99.

[10] He, Z., Wu, L., Fields, M. W., and Zhou, J. (2005) Use of microarrays with different probe sizes for monitoring gene expression.. *Appl Environ Microbiol*, **71**(9), 5154–5162.

[11] Singh-Gasson, S., Green, R. D., Yue, Y., Nelson, C., Blattner, F., Sussman, M. R., and Cerrina, F. (Oct, 1999) Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array.. *Nat Biotechnol*, **17**(10), 974–978.

[12] Matveeva, O. V., Shabalina, S. A., Nemtsov, V. A., Tsodikov, A. D., Gesteland, R. F., and Atkins, J. F. (Jul, 2003) Thermodynamic calculations and statistical correlations for oligo-probes design.. *Nucleic Acids Res*, **31**(14), 4211–4217.

[13] Held, G. A., Grinstein, G., and Tu, Y. (2003) Modeling of dna microarray data by using physical properties of hybridization.. *Proc Natl Acad Sci U S A*, **100**(13), 7575–7580.

[14] Held, G. A., Grinstein, G., and Tu, Y. (2006) Relationship between gene expression and observed intensities in dna microarrays–a modeling study.. *Nucleic Acids Res*, **34**(9), e70.

[15] Carlon, E. and Heim, T. (2006) Thermodynamics of rna/dna hybridization in high-density oligonucleotide microarrays. *Physica A: Statistical Mechanics and its Applications*, **362**(2), 433–449.

[16] Fish, D. J., Horne, M. T., Brewood, G. P., Goodarzi, J. P., Alemayehu, S., Bhandiwad, A., Searles, R. P., and Benight, A. S. (2007) Dna multiplex hybridization on microarrays and thermodynamic stability in solution: a direct comparison.. *Nucleic Acids Res*, **35**(21), 7197–7208.

[17] SantaLucia, J. (Feb, 1998) A unified view of polymer, dumbbell, and oligonucleotide dna nearest-neighbor thermodynamics.. *Proc Natl Acad Sci U S A*, **95**(4), 1460–1465.

[18] Zuker, M. (Jul, 2003) Mfold web server for nucleic acid folding and hybridization prediction.. *Nucleic Acids Res*, **31**(13), 3406–3415.

[19] Vallone, P. M. and Butler, J. M. (Aug, 2004) Autodimer: a screening tool for primer-dimer and hairpin structures.. *Biotechniques*, **37**(2), 226–231.

[20] Nielsen, H. B., Wernersson, R., and Knudsen, S. (Jul, 2003) Design of oligonucleotides for microarrays and perspectives for design of multi-transcriptome arrays.. *Nucleic Acids Res*, **31**(13), 3491–3496.

[21] Wernersson, R. and Nielsen, H. B. (Jul, 2005) Oligowiz 2.0–integrating sequence feature annotation into the design of microarray probes.. *Nucleic Acids Res*, **33**(Web Server issue), W611–W615.

[22] Wernersson, R., Juncker, A. S., and Nielsen, H. B. (2007) Probe selection for dna microarrays using oligowiz.. *Nat Protoc*, **2**(11), 2677–2691.

[23] Rouillard, J.-M., Herbert, C. J., and Zuker, M. (Mar, 2002) Oligoarray: genome-scale oligonucleotide design for microarrays.. *Bioinformatics*, **18**(3), 486–487.

[24] Rouillard, J.-M., Zuker, M., and Gulari, E. (Jun, 2003) Oligoarray 2.0: design of oligonucleotide probes for dna microarrays using a thermodynamic approach.. *Nucleic Acids Res*, **31**(12), 3057–3062.

[25] Mrowka, R., Schuchhardt, J., and Gille, C. (Dec, 2002) Oligodb–interactive design of oligo dna for transcription profiling of human genes.. *Bioinformatics*, **18**(12), 1686–1687.

[26] Nordberg, E. K. (Apr, 2005) Yoda: selecting signature oligonucleotides.. *Bioinformatics*, **21**(8), 1365–1370.

[27] Chen, H. and Sharp, B. M. (Oct, 2002) Oliz, a suite of perl scripts that assist in the design of microarrays using 50mer oligonucleotides from the 3' untranslated region.. *BMC Bioinformatics*, **3**, 27.

[28] Reymond, N., Charles, H., Duret, L., Calevro, F., Beslon, G., and Fayard, J.-M. (Jan, 2004) Roso: optimizing oligonucleotide probes for microarrays.. *Bioinformatics*, **20**(2), 271–273.

[29] Li, C. and Wong, W. The analysis of gene expression data: methods and software chapter DNA-Chip Analyzer (dChip), pp. 120–141 Springer, New York (2003).

[30] Smyth, G. K. (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments.. *Stat Appl Genet Mol Biol*, **3**, Article3.

[31] Jia, Z. and Xu, S. (2008) Bayesian mixture model analysis for detecting differentially expressed genes.. *Int J Plant Genomics*, **2008**, 892927.

[32] Akaike, H. (1974) A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, **9**, 716–723.

[33] Geisser, S. (1975) The predictive sample reuse method with application. *J Amer Stat Ass*, **70**, 320–328.

[34] Zhang, L., Miles, M. F., and Aldape, K. D. (Jul, 2003) A model of molecular interactions on short oligonucleotide microarrays.. *Nat Biotechnol*, **21**(7), 818–821.

[35] Naef, F. and Magnasco, M. O. (2003) Solving the riddle of the bright mismatches: labeling and effective binding in oligonucleotide arrays.. *Phys Rev E Stat Nonlin Soft Matter Phys*, **68**(1 Pt 1), 011906.

[36] Mei, R., Hubbell, E., Bekiranov, S., Mittmann, M., Christians, F. C., Shen, M.-M., Lu, G., Fang, J., Liu, W.-M., Ryder, T., Kaplan, P., Kulp, D., and Webster, T. A. (2003) Probe selection for high-density oligonucleotide arrays.. *Proc Natl Acad Sci U S A*, **100**(20), 11237–11242.

[37] Binder, H., Preibisch, S., and Kirsten, T. (2005) Base pair interactions and hybridization isotherms of matched and mismatched oligonucleotide probes on microarrays.. *Langmuir*, **21**(20), 9287–9302.

[38] Carlon, E., Heim, T., Wolterink, J. K., and Barkema, G. T. (Jun, 2006) Comment on ”solving the riddle of the bright mismatches: labeling and effective binding in oligonucleotide arrays”.. *Phys Rev E Stat Nonlin Soft Matter Phys*, **73**(6 Pt 1), 063901; author reply 063902.

[39] Zhang, L., Wu, C., Carta, R., and Zhao, H. (2007) Free energy of dna duplex formation on short oligonucleotide microarrays.. *Nucleic Acids Res*, **35**(3), e18.

[40] Langmuir, I. (1918) The adsorption of gases on plane surfaces of glass, mica and platinum. *J. Am. Chem. Soc.*, **40**, 1361–1403.

[41] Wick, L. M., Rouillard, J. M., Whittam, T. S., Gulari, E., Tiedje, J. M., and Hashsham, S. A. (2006) On-chip non-equilibrium dissociation curves and dissociation rate constants as methods to assess specificity of oligonucleotide probes.. *Nucleic Acids Res*, **34**(3), e26.

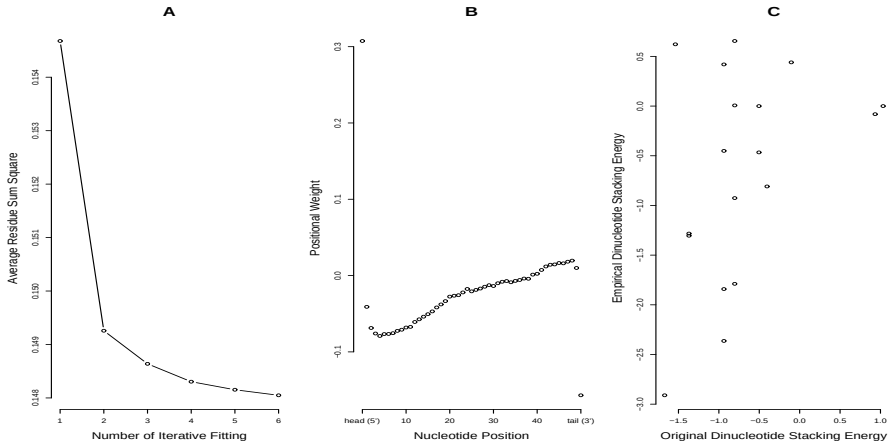


Fig. 1. ARSS, positional weights, pseudo stacking energies of PPBE model for data set 1.

A. Convergence of the PPBE model after three cycles of iterative fitting of each of positional weights and pseudo di-nucleotide stacking energies (six cycles total); **B**. Plot of positional weights; **C**. Comparison of traditional di-nucleotide stacking energies and pseudo di-nucleotide stacking energies. Y axis is the pseudo di-nucleotide stacking energies; X axis is the traditional di-nucleotide stacking energies.

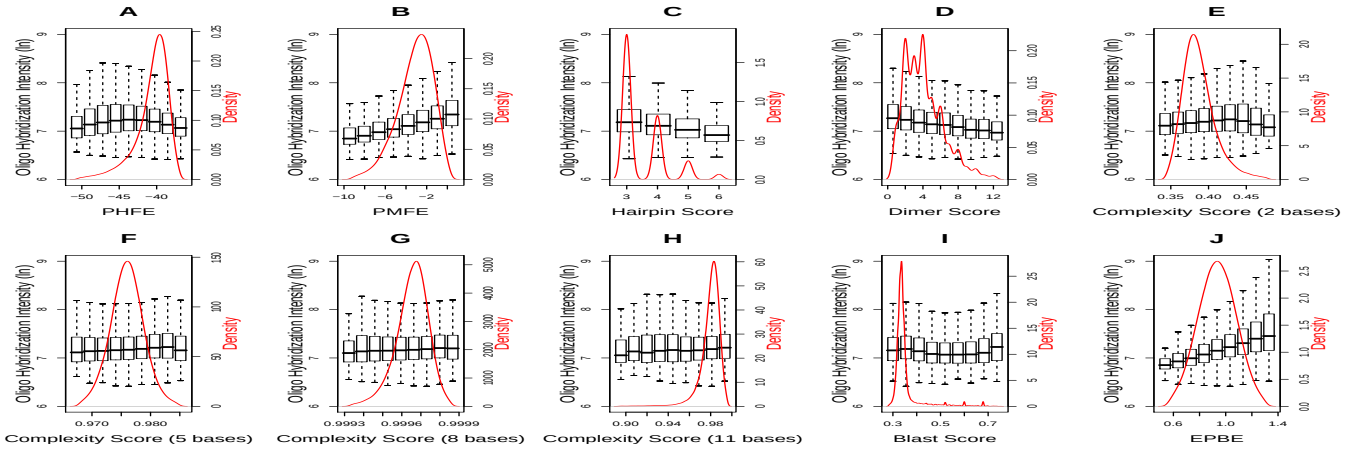


Fig. 2. Box plots show the correlation of individual probe design factors with observed oligonucleotide probe hybridization intensities for data set 1.

Density curve (red line) is computed using kernel density estimates and shows the distribution of individual probe design factors. Y axis (left) depicts probe hybridization intensity. Y axis (right) represents the density of different PDFs. X axes are: **A.** Probe hybridization free energy; **B.** Probe minimum folding energy; **C.** Probe hairpin score; **D.** Probe dimer score; **E.** Complexity score (2 bases); **F.** Complexity score (5 bases); **G.** Complexity score (8 bases); **H.** Complexity score (11 bases); **I.** Blast score; **J.** Pseudo probe binding energy.

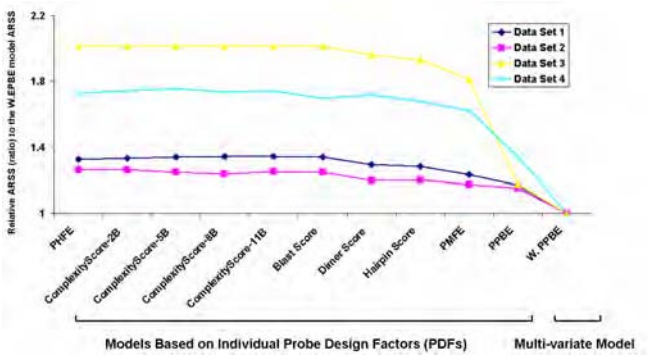


Fig. 3. Relative ARSS of different models for different data sets.

Y-axis is the ratio of each model's ARSS relative to place W. PPBE model's ARSS. From left to right, the X-axes are PHFE, Complexity Score (2 bases), Complexity Score (5 bases), Complexity Score (8 bases), Complexity Score (11 bases), blast score, dimer score, hairpin score, PMFE, PPBE, W. PPBE model.

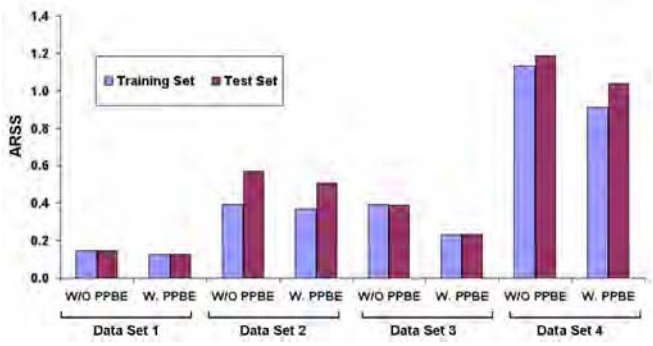


Fig. 4. Comparisons of ARSS for within-dataset validations using W/O PPBE model or W. PPBE model.

Y axis is the ARSS value. Within-dataset validation. Blue bars show the ARSS value for the training set (half of the whole data set). Brown bars show the ARSS value for the test set (half of the whole data set).

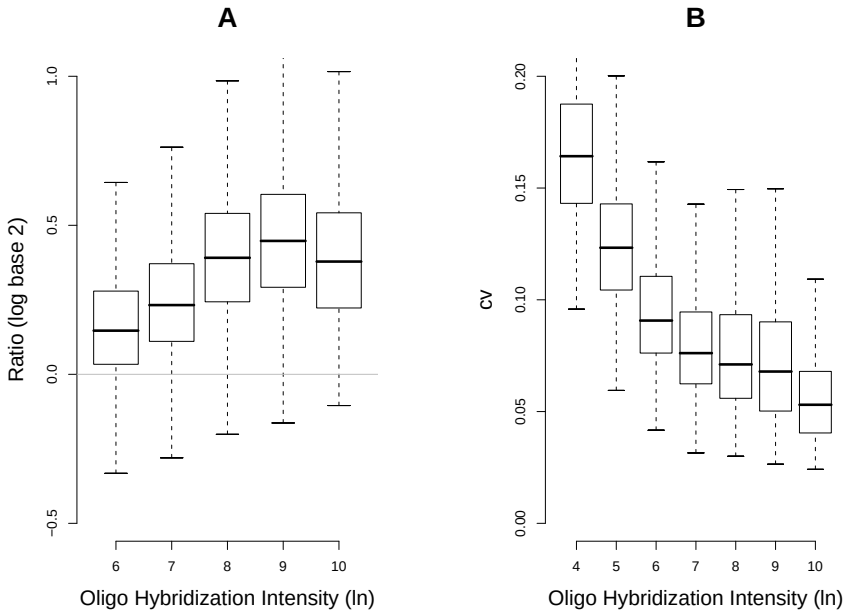


Fig. 5. Correlation of probe hybridization intensity with probe specificity and reproducibility.

A. Correlation of probe hybridization intensity with probe specificity (observed *log* base 2 transformed ratio). Grey line shows where there is no change; **B**. Correlation of oligonucleotide probe hybridization intensity with probe reproducibility, represented as coefficient of variation (cv).

Table 1. Array CGH data set used in this study.

Data Set	Microarray Platform	Sample	Manufacturer	Designer	oligos	bases	Role of data set in the analysis	sample number
1	NimbleGen HG18 whole genome CGH Array	Normal human male genomic DNA	NimbleGen Inc.	NimbleGen Inc.	137280	50	Sensitivity	6
2	NimbleGen Human Promoter Array (custom design)	Human prostate cell line (PC3M, 267B1) genomic DNA	NimbleGen Inc.	authors	220475	50	Sensitivity	4
3	NimbleGen Salmonella Whole Genome Array (custom design)	Salmonella LT2 genomic DNA	NimbleGen Inc.	authors	288238	50	Sensitivity, specificity	4
4	In-house Spotted Human Promoter Array (custom design)	Normal human lung tissue genomic DNA	authors	authors	11653	50	Sensitivity, reproducibility	205

Table 2

Simple linear model average residue square sum (ARSS) and correlation coefficients (r) for the correlation of Individual probe design factors (PDFs) with probe hybridization intensities.

	Data Set 1		Data Set 2		Data Set 3		Data Set 4	
	r	ARSS	r	ARSS	r	ARSS	r	ARSS
PHFE	0.11	0.168	0.03	0.504	0.03	0.460	0.13	1.668
PMFE	0.29	0.156	0.27	0.468	0.32	0.414	0.28	1.568
HairpinScore	0.21	0.162	0.22	0.479	0.20	0.442	0.21	1.621
DimerScore	0.19	0.164	0.23	0.478	0.17	0.448	0.15	1.660
ComplexityScore-2B	0.08	0.169	0.05	0.503	0.02	0.461	0.09	1.684
ComplexityScore-5B	0.04	0.170	0.11	0.498	0.01	0.461	0.02	1.698
ComplexityScore-8B	0.01	0.170	0.15	0.493	0.01	0.461	0.12	1.675
ComplexityScore-11B	0.01	0.170	0.10	0.498	0.02	0.461	0.10	1.683
BlastScore	0.02	0.170	0.11	0.498	0.01	0.461	0.18	1.641
EPBE	0.36	0.148	0.30	0.460	0.65	0.269	0.48	1.301

Analyzing Microarray Data Using WebArray

Wang, Yipeng [1,2]
McClelland, Michael[1]
Xia, Xiao-Qin[3,4]

[1]Department of Cancer Genetics, Sidney Kimmel Cancer Center, San Diego, CA 92121, USA
[2]Department of Pathology & Laboratory Medicine, University of California, Irvine, CA 92697, USA
[3]Genomics Core Facility, Sidney Kimmel Cancer Center, San Diego, CA 92121, USA
[4]Corresponding author (xqxia70@gmail.com)

INTRODUCTION

WebArray is a Web platform for microarray data analysis. As an analysis suite designed by bench biologists, WebArray is user-friendly for life scientists without a bioinformatics background. It is simple to use but employs powerful analysis functions. Analysis is based on files uploaded by users. For Affymetrix GeneChip data, intensity files in CEL format can be used. For two-color experiments, WebArray can recognize intensity files generated from many different software packages. WebArray provides functions for data quality control, background correction, normalization, differential analysis, and plotting on a genome map. A user-friendly aspect of WebArray is the fact that users generally do not have to change the default parameters for common experimental designs, so they are usually protected from applying the wrong statistical tools. In most cases, novice users will have no problem finding explanations for file formats or terms in the extensive help system.

RELATED INFORMATION

Supported Web browsers include Mozilla Firefox (recommended), Microsoft Internet Explorer, Opera, Flock, and Google Chrome. In WebArray's Web page, the browser window is divided into three sections: WebArray's flag is on the top panel, the left panel contains the function menu, and the rest is the work area. Generally, four steps are required to perform a new data analysis: (1) register and logon, (2) upload files, (3) select options for analysis and submit requests, and (4) browse/download results.

WebArray recognizes intensity files from many different sources, including the Affymetrix, Agilent, ArrayVision, Genepix, ImaGene, QuantArray, SMD, and SPOT software packages as well as any variable user-defined format. Only the intensity files are mandatory. Other files accepted by WebArray include the following:

- gene list file: contains a list of gene IDs and associated gene information
- target file: contains information about the samples associated with every microarray
- design file: delineates a design matrix for linear model analysis
- spot type file: identifies of different types of spots from the gene list
- genome/chromosome location file: a list of genes with information about their locations on the chromosome/genome
- composite normalization file: contains a sub-list of spots expected to be invariant between control and experiment, to be used for normalization of data between channels

Detailed descriptions can be accessed simply by clicking on the respective file-type term in the work space.

WebArray (http://www.webarray.org) was originally described by Xia et al. (2005).

METHOD

Registration and Logon

Although a guest account with full functions can be used by visitors, we encourage users to create a private account for data security. After submitting registration information, a confirmation message will be sent to the user's e-mail address. A user account will be activated immediately after the user responds to this message. Registered users can logon to WebArray with their user name and password. Passwords are encrypted for security.

1. To register:

i. Enter "http://www.webarray.org" in the address bar of the Web browser to enter WebArray's Web site.

ii. Click on the "Register" button in the function menu to enter the registration page.

iii. Enter required and (if desired) optional information, then click on the "Register" button.

iv. Check your e-mail box and follow directions in the registration confirmation message from WebArray to activate your account.

2. To log on:

i. Enter "http://www.webarray.org" in the address bar of the Web browser to enter WebArray's Web site.

ii. Enter user name/password and click on the "Sign In" button in the function menu.

iii. Click on the "WebArray" link in the function menu.

Note: The "WebArrayDB" link in the same window will take you to WebArrayDB, a database and cross-platform analysis package which will be published separately and is not part of this protocol.

File Management (Upload and Delete)

Uploaded files are stored and visible in the user's private folders. To save space on the server, users are encouraged to delete their files after all analyses have been carried out. If desired, WebArrayDB can be used for long-term storage of data in MIAME compliant formats.

3. To upload files:

i. Click on the "Upload" link in the menu.

ii. Choose/add files in the work area by clicking on the "Browse" button and selecting the respective files from your computer/network.

iii. Click on the "Upload" button on top or bottom of the work area.

JMaster's Java applet, "JumpLoader," has been integrated into WebArray as an alternative method for uploading files. Clicking on the button "Try JumpLoader" will open a file manager-like window that allows users to select local files in a drag-and-drop way. After all files are selected, click the "Start Upload" link. The uploading session will never time-out, unlike conventional HTML forms, but make sure not to close the window before all the files have uploaded successfully.

4. To delete files:

i. Click on the "Browse/Delete" link in the menu.

ii. Choose files to be deleted by clicking on the check box behind each file name.

iii. Click on the "Delete checked files" button.

Data Analysis

Users can analyze either Affymetrix GeneChip data or dual-channel data using WebArray. There are two separate dialogue frames on WebArray to deal with these two types of data. Both frames have four sections in the following order: (1) Experiment design, (2) Parameters for analysis, (3) Output options, and (4) Request name.

5. To perform data analysis:

i. Click on either the "Affymetrix" or the "Two-Color" link in the menu. A frame for data analysis will appear in the work area.

ii. Define the experimental design in the first section by selecting intensity files and defining which sample group each sample belongs to.

For Affymetrix GeneChip data, Affymetrix GeneChip CEL files (usually with ".CEL" or ".cel" as extensions of the file names) are used as intensity files. Each sample can be defined as "exp1," "exp2," "exp3," or "exp4."

For two-color data, users have to specify the correct format for the intensity files (a choice of nine different formats, including Agilent, ArrayVision, GenePix, Imagene, Quantarray, and SPOT). Channels on the arrays can subsequently be defined as "ref," "ctrl," and "exp." Note that a gene list file, or both a target file and a design file, need to be specified to enable analysis.

Important: For any experiment regardless of platform, at least two different sample groups (such as ref, ctrl, exp1, exp2, etc.) need to be present and each group must include intensity data from at least two arrays, otherwise statistical analysis will not be performed.

iii. For Affymetrix data, enter the desired comparisons. For example, "exp2-exp1; exp3-exp2" will compare (1) the difference between "exp2" and "exp1" and (2) the difference between "exp3" and "exp2." The analysis result output file will report the $\log_2$ of the ratios (i.e., exp2/exp1 and exp3/exp2) for each comparison.

iv. The second and third sections of the frame contain options for analysis and result output. The main functions that WebArray can perform include background subtraction, within-array normalization, between-array normalization, and differential statistical analysis. The default analysis parameters are suitable for the most commonly used experiment designs. In most cases, users can analyze their data without changing the settings, although more sophisticated users are free to select from any of the optional parameters to suit their specific requirements. Each analysis operation is hot-linked to a help file explaining the operation and different options in more detail.

v. In the last section, provide a name for the data analysis request.

vi. Click on the "Submit Analysis Request" button. The user will automatically be taken to a frame that displays all analysis requests submitted by that user.

Browsing Results

Submitted requests will be put in the job queue on the server. A few minutes or (occasionally) hours, depending on the level of analysis complexity and usage of the server, will be needed to complete a user request. Users do not have to wait for a request to be completed; they can close their Web browsers and return later. Results are presented in charts and tables for downloading or browsing online.

6. To browse results:

i. Follow the "Results" link in the menu. All submitted requests will be listed in the work area.

For every request, there are two links: "Browse" and "Edit." The latter brings the user to the analysis page, which facilitates changing of parameters and re-submission of jobs.

ii. Click on the "Browse" link. The work area will be redirected to a frame with all charts initially requested by the user and links to result tables. A link is offered for downloading a zip-compressed package of all results for that specific analysis request. Alternatively, users can choose to only view or download the result table, or the input parameters for that analysis request.

iii. If the user decides to view the result table, this table will be displayed. The table can be sorted in ascending or descending order for any of the column headers, including *p* value.

iv. The output data file will contain the following columns:

Columns "Block," "Row," "Column," "ID," and "Name" list the same information as in the corresponding columns in the gene list file.

"M" is the log-differential expression ratio.

"A" is the log-intensity of the spot, a measure of overall brightness of the spot.

"t" is the penalized *t*-statistic value.

"p" is the *p*-value corresponding to the *t*-statistic.

"B" is the *B* statistic; the log-odds of differential expression.

"fdr" is the estimated false discovery rate incurred by setting the threshold at the corresponding *p* value.

"fp" is the estimated number of false positives incurred by setting the threshold at the corresponding *p* value.

"fn" is the estimated number of false negatives incurred by setting threshold at the corresponding *p* value.

"M," "A," "t," "p," and "B" are calculated with linear model statistical analysis (Smyth 2004). "fdr,", "fp," and "fn" are estimated with SPLOSH (Pounds and Cheng 2004). Detailed information can be found in the WebArray help documents.

DISCUSSION

WebArray presents a simple interface for biologists to analyze microarray data. WebArray integrates functions of the LIMMA package for background correction, data normalization, and statistical analysis. More details about LIMMA can be found in the help documents of WebArray or in the literature (Smyth and Speed 2003; Smyth et al. 2005). The "affy" package (Gautier et al. 2004) is adopted for reading Affymetrix CEL files and normalizing Affymetrix gene expression data. Another independent normalization method, which is based on principal component analysis (PCA), was also included in WebArray (Stoyanova et al. 2004). The underlying algorithm for differential analysis is an eBayes-moderated *t*-test implemented in the LIMMA package (Smyth 2004), which is commonly used for conventional data from fairly simple experimental designs.

Other excellent peer Web services for microarray data analysis include SNOMAD (Colantuoni et al. 2002), ArrayQuest (Argraves et al. 2005), and GEPAS (Tárraga et al. 2008). However, WebArray has great advantages in simplicity and flexibility. The one-page analysis Web interface of WebArray makes all options clear and easier to change than the step-by-step interfaces in other software packages. A user can submit multiple analysis requests and browse the results later, which helps to save users' waiting time. Moreover, users can use WebArray just for data normalization or the integration of data from separate files.

WebArray is designed to analyze data sets from a single array platform. For complex experiments

involving more than one array platform per analysis, a more sophisticated database and analysis tool, WebArrayDB (http://www.webarraydb.org), has been deployed. Users are encouraged to first master WebArray before advancing to WebArrayDB.

ACKNOWLEDGMENTS

This work was funded in part by NIH grants R01 CA068822, U01 CA0114810, R01 AI52237, R01 AI073971, and R01 AI075093, and by a grant from the DOD, W81XWH-08-1-0720.

REFERENCES

Argraves GL, Jani S, Barth JL, Argraves WS. 2005. ArrayQuest: a web resource for the analysis of DNA microarray data. *BMC Bioinformatics* **6:** 287.

Colantuoni C, Henry G., Zeger S, Pevsner J. 2002. SNOMAD (Standardization and NOrmalization of MicroArray Data): web-accessible gene expression data analysis. *Bioinformatics* **18:** 1540-1541.

Gautier L, Cope L, Bolstad BM, Irizarry RA. 2004. affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20:** 307-315.

Pounds S, Cheng C. 2004. Improving false discovery rate estimation. *Bioinformatics* **20:** 1737-1745.

Smyth GK. 2004. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology* **3:** Iss. 1, Article 3.

Smyth GK, Speed T. 2003. Normalization of cDNA microarray data. *Methods* **31:** 265-273.

Smyth G.K., Michaud J, Scott H. 2005. The use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* **21:** 2067-2075.

Stoyanova R., Querec TD, Brown TR, Patriotis C. 2004. Normalization of single-channel DNA array data by principal component analysis. *Bioinformatics* **20:** 1772-1784.

Tárraga J, Medina I., Carbonell J, Huerta-Cepas J, Minguez P, Alloza E, Al-Shahrour F, Vegas-Azcárate S, Goetz S, Escobar P, et al. GEPAS, a web-based tool for microarray data analysis and interpretation. *Nucl Acids Res* **36(Suppl. 2):** W308-W314.

Xia X, McClelland M, Wang Y. 2005. WebArray: an online platform for microarray data analysis. *BMC Bioinformatics* **6:** 306.

Electronic Acknowledgement Receipt

EFS ID:	5511266
Application Number:	
International Application Number:	PCT/US09/47285
Confirmation Number:	1680
Title of Invention:	METHODS TO TREAT SOLID TUMORS
First Named Inventor/Applicant Name:	Vivocure, Inc.
Customer Number:	69403
Correspondence Address:	Bruce D. Grant Grant Anderson LLP c/o PortfolioIP P.O. Box 52050 - Minneapolis MN 55402 US (858) 623-3226 bgrant@granllp.com
Filer:	Bruce David Grant/Dave Glisson
Filer Authorized By:	Bruce David Grant
Attorney Docket Number:	VIV-1001-PC
Receipt Date:	12-JUN-2009
Filing Date:	
Time Stamp:	21:03:56
Application Type:	International Application for filing in the US receiving office

Payment information:

Submitted with Payment	yes

Payment Type	Credit Card
Payment was successfully received in RAM	$3400
RAM confirmation Number	5121
Deposit Account	503473
Authorized User	GRANT,BRUCE

The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:

Charge any Additional Fees required under 37 CFR 1.445 (International application filing, processing and search fees)

Charge any Additional Fees required under 37 CFR 1.17(t) (Acceptance of an unintentionally delayed claim for priority)

Charge any Additional Fees required under PCT Rule 14

Charge any Additional Fees required under PCT Rule 15

Charge any Additional Fees required under PCT Rule 16

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1		VIV-1001-PC_PCT_Request.pdf	672544 520ad2cb70a29d760626d2188b9a654b66dcfc4f	yes	6

Multipart Description/PDF files in .zip description		
Document Description	**Start**	**End**
RO/101 - Request form for new IA - Conventional	1	5
RO/101 - Annex (fee calculation sheet)	6	6

Warnings:

Information:

2		VIV-1001-PC_SPEC_2009-06-12.pdf	698462 f2d85da6fdcb373937783093c20c025f00802dd3	yes	135

Multipart Description/PDF files in .zip description		
Document Description	**Start**	**End**
Specification	1	129
Claims	130	134
Abstract	135	135

Warnings:

Information:

3	Drawings-only black and white line drawings	VIV-1001-PC_DRAW_2009-06-12.pdf	734370 61741e080f2978345638ad6a6375feb4cfe2732e	no	2

Warnings:

Information:					
4	Fee Worksheet (PTO-875)	fee-info.pdf	36297 f649026727286f21b6155ffef34c5b9bce1761ba	no	2
Warnings:					
Information:					
			Total Files Size (in bytes):	2141673	

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Electronic Patent Application Fee Transmittal

Application Number:	
Filing Date:	
Title of Invention:	METHODS TO TREAT SOLID TUMORS
First Named Inventor/Applicant Name:	Vivocure, Inc.
Filer:	Bruce David Grant/Dave Glisson
Attorney Docket Number:	VIV-1001-PC

International Application for filing in the US receiving office Filing Fees

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
Transmittal fee	1601	1	240	240
Intl Filing Fee (1st-30 Pgs.) PCT Easy	1701	1	1095	1095
Suppl. Intl Filing Fee (each page > 30)	1703	112	13	1456
International Search (KIPO)	1709	1	609	609
Pages:				
Claims:				
Miscellaneous-Filing:				
Petition:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				
Miscellaneous:				
		Total in USD ($)		**3400**

PCT

REQUEST

The undersigned requests that the present international application be processed according to the Patent Cooperation Treaty.

For receiving Office use only

International Application No.

International Filing Date

Name of receiving Office and "PCT International Application"

Applicant's or agent's file reference *(if desired) (12 characters maximum)* VIV-1001-PC

Box No. I TITLE OF INVENTION

METHODS TO TREAT SOLID TUMORS

Box No. II APPLICANT ☐ This person is also inventor

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

VIVOCURE, INC.
1932 Burton Street
San Diego, California 92111
UNITED STATES OF AMERICA

Telephone No.

Facsimile No.

Applicant's registration No. with the Office

☐ **E-mail authorization**: Marking this check-box authorizes the receiving Office, the International Searching Authority, the International Bureau and the International Preliminary Examining Authority to use the e-mail address indicated in this Box to send, if the Office or Authority so wishes, advance copies of notifications in respect of this international application. *(See also the Notes to Boxes Nos. II and III.)*

E-mail address

State *(that is, country)* of nationality: US

State *(that is, country)* of residence: US

This person is applicant for the purposes of: ☐ all designated States ☒ all designated States except the United States of America ☐ the United States of America only ☐ the States indicated in the Supplemental Box

Box No. III FURTHER APPLICANT(S) AND/OR (FURTHER) INVENTOR(S)

☒ Further applicants and/or (further) inventors are indicated on a continuation sheet.

Box No. IV AGENT OR COMMON REPRESENTATIVE; OR ADDRESS FOR CORRESPONDENCE

The person identified below is hereby/has been appointed to act on behalf of the applicant(s) before the competent International Authorities as: ☒ agent ☐ common representative

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country.)*

GRANT, Bruce D.
Grant Anderson LLP
c/o PortfolioIP
P.O. Box 52050
Minneapolis, Minnesota 55402
UNITED STATES OF AMERICA

Telephone No.
(858) 623-3226

Facsimile No.
(612) 332-8352

Agent's registration No. with the Office
47,608

☐ **E-mail authorization**: Marking this check-box authorizes the receiving Office, the International Searching Authority, the International Bureau and the International Preliminary Examining Authority to use the e-mail address indicated in this Box to send, if the Office or Authority so wishes, advance copies of notifications in respect of this international application. *(See also the Notes to Boxes Nos. II and III.)*

E-mail address

☐ **Address for correspondence**: Mark this check-box where no agent or common representative is/has been appointed and the space above is used instead to indicate a special address to which correspondence should be sent.

Box No. III FURTHER APPLICANT(S) AND/OR (FURTHER) INVENTOR(S)

If none of the following sub-boxes is used, this sheet should not be included in the request.

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

ARRACH, Nabil
1932 Burton Street
San Diego, California 92111
UNITED STATES OF AMERICA

This person is:
- ☐ applicant only
- ☒ applicant and inventor
- ☐ inventor only *(If this check-box is marked, do not fill in below.)*

Applicant's registration No. with the Office

State *(that is, country)* of nationality: US

State *(that is, country)* of residence: US

This person is applicant for the purposes of: ☐ all designated States ☐ all designated States except the United States of America ☒ the United States of America only ☐ the States indicated in the Supplemental Box

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

MCCLELLAND, Michael
1932 Burton Street
San Diego, California 92111
UNITED STATES OF AMERICA

This person is:
- ☐ applicant only
- ☒ applicant and inventor
- ☐ inventor only *(If this check-box is marked, do not fill in below.)*

Applicant's registration No. with the Office

State *(that is, country)* of nationality: US

State *(that is, country)* of residence: US

This person is applicant for the purposes of: ☐ all designated States ☐ all designated States except the United States of America ☒ the United States of America only ☐ the States indicated in the Supplemental Box

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

This person is:
- ☐ applicant only
- ☐ applicant and inventor
- ☐ inventor only *(If this check-box is marked, do not fill in below.)*

Applicant's registration No. with the Office

State *(that is, country)* of nationality:

State *(that is, country)* of residence:

This person is applicant for the purposes of: ☐ all designated States ☐ all designated States except the United States of America ☐ the United States of America only ☐ the States indicated in the Supplemental Box

Name and address: *(Family name followed by given name; for a legal entity, full official designation. The address must include postal code and name of country. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below.)*

This person is:
- ☐ applicant only
- ☐ applicant and inventor
- ☐ inventor only *(If this check-box is marked, do not fill in below.)*

Applicant's registration No. with the Office

State *(that is, country)* of nationality:

State *(that is, country)* of residence:

This person is applicant for the purposes of: ☐ all designated States ☐ all designated States except the United States of America ☐ the United States of America only ☐ the States indicated in the Supplemental Box

☐ Further applicants and/or (further) inventors are indicated on another continuation sheet.

Supplemental Box *If the Supplemental Box is not used, this sheet should not be included in the request.*

1. *If, in any of the Boxes, except Boxes Nos. VIII(i) to (v) for which a special continuation box is provided, **the space is insufficient** to furnish all the information: in such case, write "Continuation of Box No...." (indicate the number of the Box) and furnish the information in the same manner as required according to the captions of the Box in which the space was insufficient, in particular:*

   (i) ***if more than one person is to be indicated as applicant and/or inventor** and no "continuation sheet" is available: in such case, write "Continuation of Box No. III" and indicate for each additional person the same type of information as required in Box No. III. The country of the address indicated in this Box is the applicant's State (that is, country) of residence if no State of residence is indicated below;*

   (ii) *if, in Box No. II or in any of the sub-boxes of Box No. III, the indication **"the States indicated in the Supplemental Box"** is checked: in such case, write "Continuation of Box No. II" or "Continuation of Box No. III" or "Continuation of Boxes No. II and No. III" (as the case may be), indicate the name of the applicant(s) involved and, next to (each) such name, the State(s) (and/or, where applicable, ARIPO, Eurasian, European or OAPI patent) for the purposes of which the named person is applicant;*

   (iii) *if, in Box No. II or in any of the sub-boxes of Box No. III, **the inventor or the inventor/applicant is not inventor for the purposes of all designated States or for the purposes of the United States of America**: in such case, write "Continuation of Box No. II" or "Continuation of Box No. III" or "Continuation of Boxes No. II and No. III" (as the case may be), indicate the name of the inventor(s) and, next to (each) such name, the State(s) (and/or, where applicable, ARIPO, Eurasian, European or OAPI patent) for the purposes of which the named person is inventor;*

   (iv) *if, in addition to the agent(s) indicated in Box No. IV, there are **further agents**: in such case, write "Continuation of Box No. IV" and indicate for each further agent the same type of information as required in Box No. IV;*

   (v) *if, in Box No. VI, there are **more than four earlier applications whose priority is claimed**: in such case, write "Continuation of Box No. VI" and indicate for each additional earlier application the same type of information as required in Box No. VI.*

2. *If the applicant intends to make an indication of the wish that the international application be treated, in certain designated States, as an application for a patent of addition, certificate of addition, inventor's certificate of addition or utility certificate of addition: in such a case, write the name or two-letter code of each designated State concerned and the indication "**patent of addition,**" "**certificate of addition,**" "**inventor's certificate of addition**" or "**utility certificate of addition,**" the number of the parent application or parent patent or other parent grant and the date of grant of the parent patent or other parent grant or the date of filing of the parent application (Rules 4.11(a)(i) and 49bis.1(a) or (b)).*

3. *If the applicant intends to make an indication of the wish that the international application be treated, in the United States of America, as a continuation or continuation-in-part of an earlier application: in such a case, write "United States of America" or "US" and the indication "**continuation**" or "**continuation-in-part**" and the number and the filing date of the parent application (Rules 4.11(a)(ii) and 49bis.1(d)).*

Continuation of Box IV:

William B. Anderson, Registration No. 41,585
Sheryl R. Silverstein, Registration No. 40,812
Tobey M. Tam, Registration No. 54,484

Box No. V DESIGNATIONS

The filing of this request **constitutes under Rule 4.9(a) the designation** of all Contracting States bound by the PCT on the international filing date, for the grant of every kind of protection available and, where applicable, for the grant of both regional and national patents.

However,

☐ DE Germany **is not designated** for any kind of national protection

☐ JP Japan **is not designated** for any kind of national protection

☐ KR Republic of Korea **is not designated** for any kind of national protection

☐ RU Russian Federation **is not designated** for any kind of national protection

(The check-boxes above may only be used to exclude (irrevocably) the designations concerned if, at the time of filing or subsequently under Rule 26bis.1, the international application contains in Box No. VI a priority claim to an earlier national application filed in the particular State concerned, in order to avoid the ceasing of the effect, under the national law, of this earlier national application.)

Box No. VI PRIORITY CLAIM

The priority of the following earlier application(s) is hereby claimed:

Filing date of earlier application *(day/month/year)*	Number of earlier application	Where earlier application is:		
		national application: country or Member of WTO	regional application: regional Office	international application: receiving Office
item (1) 13 June 2008 (13.06.2008)	61/061,576	US		
item (2)				
item (3)				
item (4)				

☐ Further priority claims are indicated in the Supplemental Box.

Transmit certified copy: the receiving Office is requested to prepare and transmit to the International Bureau a certified copy of the earlier application(s) *(only if the earlier application was filed with the Office which for the purposes of this international application is the receiving Office)* identified above as:

☒ all items ☐ item (1) ☐ item (2) ☐ item (3) ☐ item (4) ☐ other, see Supplemental Box

Restore the right of priority: the receiving Office is requested to restore the right of priority for the earlier application(s) identified above or in the Supplemental Box as item(s) (_____). *(See also the Notes to Box No. VI; further information **must** be provided to support a request to restore the right of priority.)*

Incorporation by reference: where an element of the international application referred to in Article 11(1)(iii)(d) or (e) or a part of the description, claims or drawings referred to in Rule 20.5(a) is not otherwise contained in this international application but is completely contained in an earlier application whose priority is claimed on the date on which one or more elements referred to in Article 11(1)(iii) were first received by the receiving Office, that element or part is, subject to confirmation under Rule 20.6, incorporated by reference in this international application for the purposes of Rule 20.6.

Box No. VII INTERNATIONAL SEARCHING AUTHORITY

Choice of International Searching Authority (ISA) *(if more than one International Searching Authority is competent to carry out the international search, indicate the Authority chosen; the two-letter code may be used):*

ISA/ KR

Box No. IX CHECK LIST; LANGUAGE OF FILING

This international application **contains**:

(a) **on paper**, the following number of sheets:

request (including declaration and supplemental sheets)	:	5
description (excluding sequence listing and/or tables related thereto)	:	129
claims	:	5
abstract	:	1
drawings	:	2
Sub-total number of sheets	:	142
sequence listing	:	
tables related thereto	:	
(for both, actual number of sheets if filed on paper, whether or not also filed in electronic form; see (c) below)		
Total number of sheets	:	142

(b) ☐ **only in electronic form** (Section 801(a)(i))
- (i) ☐ sequence listing
- (ii) ☐ tables related thereto

(c) ☐ **also in electronic form** (Section 801(a)(ii))
- (i) ☐ sequence listing
- (ii) ☐ tables related thereto

Type and number of carriers (diskette, CD-ROM, CD-R or other) on which are contained the
- ☐ sequence listing: . . . . . .
- ☐ tables related thereto: . . . . . .

(additional copies to be indicated under items 9(ii) and/or 10(ii), in right column)

This international application is **accompanied** by the following item(s) *(mark the applicable check-boxes below and indicate in right column the number of each item)*:

		Number of items
1. ☒ fee calculation sheet	:	1
2. ☐ original separate power of attorney	:	
3. ☐ original general power of attorney	:	
4. ☐ copy of general power of attorney; reference number, if any:	:	
5. ☐ statement explaining lack of signature	:	
6. ☐ priority document(s) identified in Box No. VI as item(s):	:	
7. ☐ translation of international application into *(language)*:	:	
8. ☐ separate indications concerning deposited microorganism or other biological material	:	
9. ☐ sequence listing in electronic form *(indicate type and number of carriers)*		
(i) ☐ copy submitted for the purposes of international search under Rule 13*ter* only (and not as part of the international application)	:	
(ii) ☐ *(only where check-box (b)(i) or (c)(i) is marked in left column)* additional copies including, where applicable, the copy for the purposes of international search under Rule 13*ter*	:	
(iii) ☐ together with relevant statement as to the identity of the copy or copies with the sequence listing mentioned in left column	:	
10. ☐ tables in electronic form related to sequence listing *(indicate type and number of carriers)*		
(i) ☐ copy submitted for the purposes of international search under Section 802(b-*quater*) only (and not as part of the international application)	:	
(ii) ☐ *(only where check-box (b)(ii) or (c)(ii) is marked in left column)* additional copies including, where applicable, the copy for the purposes of international search under Section 802(b-*quater*)	:	
(iii) ☐ together with relevant statement as to the identity of the copy or copies with the tables mentioned in left column	:	
11. ☐ copy of results of earlier search(es) (Rule 12*bis*.1(a))	:	
12. ☐ other *(specify)*:	:	

Figure of the drawings which should accompany the abstract: none

Language of filing of the international application: English

Box No. X SIGNATURE OF APPLICANT, AGENT OR COMMON REPRESENTATIVE

Next to each signature, indicate the name of the person signing and the capacity in which the person signs (if such capacity is not obvious from reading the request).

/Bruce Grant/

Bruce D. Grant
Registration No. 47,608

For receiving Office use only

1. Date of actual receipt of the purported international application:	2. Drawings: ☐ received:
3. Corrected date of actual receipt due to later but timely received papers or drawings completing the purported international application:	
4. Date of timely receipt of the required corrections under PCT Article 11(2):	☐ not received:
5. International Searching Authority (if two or more are competent): ISA /	6. ☐ Transmittal of search copy delayed until search fee is paid

For International Bureau use only

Date of receipt of the record copy by the International Bureau:

This sheet is not part of and does not count as a sheet of the international application.

PCT

FEE CALCULATION SHEET
Annex to the Request

For receiving Office use only

International Application No.

Date stamp of the receiving Office

Applicant's or agent's file reference: VIV-1001-PC

Applicant: VIVOCURE, INC.

CALCULATION OF PRESCRIBED FEES

1. TRANSMITTAL FEE . . . . . 240.00 T

2. SEARCH FEE . . . . . 609.00 S

International search to be carried out by KR

(If two or more International Searching Authorities are competent to carry out the international search, indicate the name of the Authority which is chosen to carry out the international search.)

3. INTERNATIONAL FILING FEE

Where items (b) and/or (c) of Box No. IX apply, enter **Sub-total number of sheets** } 142
Where items (b) and (c) of Box No. IX do not apply, enter **Total number of sheets** }

i1 first 30 sheets 1095.00 i1

i2 112 (number of sheets in excess of 30) x 13.00 (fee per sheet) = 1456.00 i2

i3 additional component (only if a sequence listing and/or tables related thereto are filed in electronic form under Section 801(a)(i), or both in that form and on paper, under Section 801(a)(ii)):

400 x _____ (fee per sheet) = _____ i3

Add amounts entered at i1, i2 and i3 and enter total at I . . . . 2551.00 I

(Applicants from certain States are entitled to a reduction of 90% of the international filing fee. Where the applicant is (or all applicants are) so entitled, the total to be entered at I is 10% of the international filing fee.)

4. FEE FOR PRIORITY DOCUMENT *(if applicable)* . . . . . P

5. FEE FOR RESTORATION OF THE RIGHT OF PRIORITY *(if applicable)* . RP

6. FEE FOR EARLIER SEARCH DOCUMENTS *(if applicable)* . . . . . ES

7. TOTAL FEES PAYABLE $3400.00

Add amounts entered at T, S, I, P, RP and ES, and enter total in the TOTAL box — TOTAL

MODE OF PAYMENT *(Not all modes of payment may be available at all receiving Offices)*

- ☐ authorization to charge deposit account (see below)
- ☐ postal money order
- ☐ cash
- ☐ coupons
- ☐ cheque
- ☐ bank draft
- ☐ revenue stamps
- ☒ other *(specify)*: credit card

AUTHORIZATION TO CHARGE (OR CREDIT) DEPOSIT ACCOUNT
(This mode of payment may not be available at all receiving Offices)

- ☐ Authorization to charge the total fees indicated above.
- ☒ *(This check-box may be marked only if the conditions for deposit accounts of the receiving Office so permit)* Authorization to charge any deficiency or credit any overpayment in the total fees indicated above.
- ☐ Authorization to charge the fee for priority document.

Receiving Office: RO/ US

Deposit Account No.: 50-3473

Date: 12 June 2009

Name: Bruce D. Grant

Signature: /Bruce Grant/

See Notes to the fee calculation sheet

PATENT
VIV-1001-PC

METHODS TO TREAT SOLID TUMORS

Related Patent Application(s)

This application claims the benefit of U.S. provisional patent application no. 61/061,576 filed on June 13, 2008, entitled "Method to Treat Solid Tumors, and designated by Attorney Docket number 655233000100. The entire content of the foregoing patent application is incorporated herein by reference, including, without limitation, all text, tables and drawings.

Statement of Government Support

This invention was made in part with government support under Grant Nos. R01 AI034829, R01 AI052237, and R21 AI057733 awarded by the National Institutes of Health (NIH) and Grant Nos. TRDRP 16KT-0045 to Sidney Kimmel Cancer Center from the Tobacco-Related Disease Research Program of California and grants CA 103563; CA 119811 and DCD grant W81XWH-06-0117 to AntiCancer. The government has certain rights in this invention.

Field of the Invention

The invention relates in part to compositions and methods selectively to target solid tumors. More specifically, it concerns compositions comprising expression systems for cytotoxic proteins under the control of promoters active in tumors.

Background

A wide range of bacteria (e.g., *Escherichia*, *Salmonella*, *Clostridium, Listeria,* and *Bifidobacterium*, for example) have been shown to preferentially colonize solid tumors. *Salmonella enterica* and avirulent derivatives may effect some degree of tumor reduction by the presence of the bacteria in the solid tumor. The internal environment of solid tumors is not well understood and may present favorable growing conditions to colonizing bacteria.

Summary

The environment inside solid tumors is very different from that in normal, healthy tissue. Solid tumors often are poorly vascularized and sometimes have areas of necrosis. The poor

vascularization contributes to hypoxic or anoxic areas that can extend to about 100 micrometers from the vasculature of the solid tumor. Solid tumors also can have an internal pH lower than the organism's normal pH. Necrosis in solid tumors can lead to a nutrient rich environment where bacteria capable of growing in low oxygen conditions can flourish. In addition to the nutrient rich environment, the internal spaces of solid tumors also offer some degree of protection from a host organisms' immune system, and thus shield the bacteria from the hosts' immune response. These conditions may cause bacteria to express genes that are not normally expressed in normal, healthy tissues. These factors may contribute to the preferential colonization of solid tumors as compared to other normal tissue.

The internal environment of tumors may offer regulatory conditions not well understood, in addition to low oxygen and low pH. Promoters are nucleotide sequences that in part regulate the production of mRNA from coding sequences in genomic DNA. The mRNA then can be translated into a polypeptide having a particular biological activity. Bacterial promoters that are preferentially activated in tumors have been identified by methods described herein, and compositions that contain such promoters, and methods for using them, also are described.

Thus, provided herein are isolated nucleic acid molecules that comprise a recombinant expression system, which expression system comprises a nucleotide sequence encoding a toxic or therapeutic RNA (e.g., mRNA, tRNA, rRNA, siRNA, ribozyme, and the like), a protein or an RNA or protein that participates in generating a toxin or therapeutic agent, or a nucleotide sequence encoding a toxic or therapeutic agent, RNA or protein which can mobilize the subjects immune response, operably linked to a heterologous promoter which promoter is preferentially activated in solid tumors. In certain embodiments, the heterologous promoter sequence can be a naturally occurring promoter sequence. In some embodiments the promoter can be an *Enterobacteriaceae* promoter, and in certain embodiments the promoter is a *Salmonella* promoter. In some embodiments, the promoter may comprise (i) a nucleotide sequence of Table 2A, (ii) a functional promoter nucleotide sequence 80% or more identical to a nucleotide sequence of Table 2A, or (iii) or a functional promoter subsequence of (i) or (ii). In certain embodiments, the functional promoter subsequence is about 20 to about 150 nucleotides in length.

The term "preferentially activated in solid tumors" as used herein refers to a nucleotide sequence that expresses a polypeptide from a coding sequence in tumors at a level of at least two-fold more than the same polypeptide from the same coding sequence is expressed in non-tumor cells. The

polypeptide may be expressed at detectable levels in non-tumor cells or tissue in some embodiments, and in certain embodiments, the polypeptide is not detectably expressed in non-tumor cells or tissue. As an example, preferential activation can be determined using (i) cells from the spleen as non-tumor cells and (ii) PC3 prostate cancer cells in a tumor xenograft for tumor cells. A reference level of the amount of polypeptide produced can be determined by the promoter expression in the bacterial culture samples, before injecting aliquots of the sample into mice (e.g., measuring GFP expression in the overnight cultures prepared to inject mice, also known as the input library). In some embodiments, preferential activation in solid tumors is identified by utilizing spleen, PC3 tumor xenograft and reference level (i.e., input) determinations described in Example 2 hereafter. In certain embodiments, a promoter is preferentially activated in a tumor of a living organism. In some embodiments, there can be two references used on the arrays described in Examples 1 and 2. One reference can be a library of all plasmids extracted from bacteria grown overnight in LB+Amp (see below) culture broth, as described above. Another suitable reference that can be used would be to compare the profile of bacteria expressing GFP from a particular tissue of interest to the profile of all bacteria (e.g., GFP expresser and non-expressers, for example) isolated from the same tissue of interest.

Also provided are suitable delivery vectors for administering the isolated nucleic acid which may comprise a recombinant expression system. In some embodiments, recombinant host cells that contain the nucleic acid molecules described above or below may be used to delivery the expression system to a patient or subject. In certain embodiments, the cells may be avirulent *Salmonella* cells. Also provided are pharmaceutical compositions which can comprise the nucleic acid reagents isolated, generated or modified by methods described herein, or cells which harbor such nucleic acid reagents.

Also provided, in certain embodiments, are methods to treat solid tumors, which methods can comprise administering to a subject harboring a tumor the nucleic acid molecules isolated or generated as described herein, the cells containing them or compositions comprising the nucleic acid reagents and/or cells harboring them.

Also provided, in some embodiments, are methods for identifying a promoter preferentially activated in tumor tissue which method comprises: (a) providing a library of expression systems each may comprise a nucleotide sequence encoding a detectable protein operably linked to a different candidate promoter; (b) providing the library to solid tumor tissue and to normal tissue; (c)

identifying cells from each tissue that show high levels of expression of the detectable protein; and (d) obtaining the expressions systems from the cells that produce greater levels of detectable protein in tumor tissue as compared to normal tissue, and identifying the promoters of the expression system. In some embodiments, the method may further comprise scoring the promoters identified in (d) (e.g., described below in Example 2). In some embodiments, the library is provided in recombinant host cells. In certain embodiments, the library of DNA fragments can be a random set of fragments from a bacterial genome (e.g., *Salmonella* genome, for example) in the range of about 25 to about 10,000 base pairs (bp) in length, for example. In some embodiments, the library may comprise known nucleic acid regions or known promoter regions from a bacterial genome in the range of about 25 to about 10,000 bp in length, for example.

In certain embodiments, the promoters can be *Salmonella* promoters and the recombinant host cells can be *Salmonella*. In some embodiments, the candidate promoters are from bacteria, or are 80% or more identical to promoters from bacteria. In certain embodiments, the bacteria can be *Enterobacteriaceae*, and in some embodiments the *Enterobacteriaceae* can be *Salmonella*. Also provided, in some embodiments, is an expression system which comprises a nucleotide sequence encoding a toxic or therapeutic RNA or protein or an RNA or protein that participates in generating a desired toxin or therapeutic agent operably linked to a promoter identified by the methods described herein. Also provided herein, in certain embodiments, are recombinant host cells that may comprise an expression system described herein.

Also provided, in certain embodiments, are methods to treat solid tumors which methods comprise administering an expression system described herein or cells containing an expression system described herein, to a subject harboring a solid tumor.

Also provided, in some embodiments, is an expression system which may comprise a first promoter nucleotide sequence operably linked to a first coding sequence and second promoter nucleotide sequence operably linked to a second coding sequence, where: the first coding sequence and the second coding sequence encode polypeptides that individually do not inhibit tumor growth; polypeptides encoded by the first coding sequence and the second coding sequence, in combination, inhibit tumor growth; and the first promoter nucleotide sequence and the second promoter nucleotide sequence can be preferentially activated in solid tumors of living organisms. In certain embodiments, one or more of the promoter nucleotide sequences can be preferentially activated in solid tumors (e.g., one promoter is constitutive and one promoter is

preferentially activated in solid tumors). In some embodiments, the first promoter nucleotide sequence and the second promoter nucleotide sequence can be in the same nucleic acid molecule. In certain embodiments, the first promoter nucleotide sequence and the second promoter nucleotide sequence may be in different nucleic acid molecules. In some embodiments, the first promoter nucleotide sequence and the second promoter nucleotide sequence can be bacterial nucleotide sequences. In certain embodiments, the bacterial sequences may be *Enterobacteriaceae* sequences, and in some embodiments the *Enterobacteriaceae* sequences can be *Salmonella* sequences. In certain embodiments, the different nucleic acid molecules can be disposed in the same recombinant host cell, and in some embodiments, the different nucleic acid molecules can be disposed in different recombinant host cells of the same species. In some embodiments, the different recombinant host cells can be different bacterial species.

In some embodiments, expression systems as described herein can produce two components that interact to provide a functional therapeutic agent, where: a first coding sequence may encode an enzyme, a second coding sequence may encode a prodrug, and the enzyme can process the prodrug into a drug that inhibits tumor growth. In certain embodiments, expression systems as described herein can produce two components that interact to provide a functional therapeutic agent, where; the first coding sequence may encode a first polypeptide, the second coding sequence can encode a second polypeptide, and the first polypeptide and the second polypeptide can form a complex that inhibits tumor growth.

In some embodiments, the first promoter nucleotide sequence, the second promoter nucleotide sequence, or the first promoter nucleotide sequence and the second promoter nucleotide sequence can comprise (i) a nucleotide sequence of Table 2A, (ii) a functional promoter nucleotide sequence 80% or more identical to a nucleotide sequence of Table 2A, or (iii) or a functional promoter subsequence of (i) or (ii). In certain embodiments, the functional promoter subsequence is about 20 to about 150 nucleotides in length. In some embodiments, expression systems described herein may be contained in recombinant host cells, and in certain embodiments, the recombinant host cells can be avirulent *Salmonella*.

Also provided, in certain embodiments, is an expression system which comprises three or more promoters operably linked to three or more coding sequences, where one, two, or more of the promoter nucleotide sequences are preferentially activated in solid tumors. In some embodiments,

the coding sequences encode polypeptides that individually do not inhibit tumor growth and polypeptides encoded by the coding sequences, in combination, inhibit tumor growth.

Certain embodiments are described further in the following description, examples, claims and drawings.

Brief Description of the Drawings

The drawings illustrate embodiments of the invention and are not limiting. For clarity and ease of illustration, the drawings are not made to scale and, in some instances, various aspects may be shown exaggerated or enlarged to facilitate an understanding of particular embodiments.

FIG. 1 is a flow diagram illustrating the procedure used to construct the nucleic acid libraries used to identify and isolate *Salmonella* genomic sequences corresponding to promoter elements. FIG. 2 shows photographs taken of tumors expressing GFP, demonstrating the *in vivo* function of the promoter elements identified and isolated using the methods described herein.

Detailed Description

Methods and compositions described herein have been designed to identify and isolate nucleic acid promoter sequences that can be preferentially activated under unique conditions found inside solid tumors of living organisms. Without being limited by any particular theory or to any particular class of inducible promoters, promoter identification methods described herein may be utilized to identify all classes of promoters that are preferentially active in solid tumors of living organisms. In some embodiments, promoter identification methods described herein can potentially identify promoters activated by the following classes of regulatory agents, including but not limited to, gases (e.g., oxygen, nitrogen, carbon dioxide and the like), pH (e.g., acidic pH or basic pH), metals (e.g., iron, copper and the like), hormones (e.g., steroids, peptides and the like), and various cellular components (e.g., purines, pyrimidines, sugars, and the like). The methods and compositions described herein also can be used to identify promoters preferentially active in any part of the body of a living organism, including wounds or diseased parts of the body, for example. Non-limiting examples of solid tumors that may be treated by methods and compositions described herein are sarcomas (e.g., rhabdomyosarcoma, osteosarcoma, and the like, for example), lymphomas, blastomas (e.g., hepatocblastoma, retinoblastoma, and neuroblastom, for example),

germ cell tumors (e.g., choriocarcinoma, and endodermal sinus tumor, for example), endocrine tumors, and carcinomas (e.g., adrenocortical carcinoma, colorectal carcinoma, hepatocellular carcinoma, for example).

Promoter elements preferentially activated in solid tumors of living organisms, identified and isolated using the methods described herein, can be used in targeted, tumor specific therapies. In some embodiments a promoter nucleotide sequence (e.g., heterologous promoter) is operably linked to a nucleotide sequence encoding one or more therapeutic agents. In some embodiments, the promoter sequence can be a naturally occurring nucleic acid sequence. A therapeutic agent includes, without limitation, a toxin (e.g.,ricin, diphtheria toxin, abrin, and the like), a peptide, polypeptide or protein with therapeutic activity (e.g., methioninase, nitroreductase, antibody, antibody fragment, single chain antibody), a prodrug (e.g., CB1954), an RNA molecule (e.g., siRNA, ribozyme and the like, for example). The structures of such therapeutic agents are known and can be adapted to systems described herein, and can be from any suitable organism, such as a prokaryote (e.g., bacteria) or eukaryote (e.g., yeast, fungi, reptile, avian, mammal (e.g., human or non-human)), for example.

Antibodies sometimes are IgG, IgM, IgA, IgE, or an isotype thereof (e.g., IgG1, IgG2a, IgG2b or IgG3), sometimes are polyclonal or monoclonal, and sometimes are chimeric, humanized or bispecific versions of such antibodies. Polyclonal and monoclonal antibodies that bind specific antigens are commercially available, and methods for generating such antibodies are known. In general, polyclonal antibodies are produced by injecting an isolated antigen into a suitable animal (e.g., a goat or rabbit); collecting blood and/or other tissues from the animal containing antibodies specific for the antigen and purifying the antibody. Methods for generating monoclonal antibodies, in general, include injecting an animal with an isolated antigen (e.g., often a mouse or a rat); isolating splenocytes from the animal; fusing the splenocytes with myeloma cells to form hybridomas; isolating the hybridomas and selecting hybridomas that produce monoclonal antibodies which specifically bind the antigen (e.g., Kohler & Milstein, Nature 256:495 497 (1975) and StGroth & Scheidegger, J Immunol Methods 5:1 21 (1980)). Examples of monoclonal antibodies are anti MDM 2 antibodies, anti-p53 antibodies (pAB421, DO 1, and an antibody that binds phosphoryl-ser15), anti-dsDNA antibodies and anti-BrdU antibodies, are described hereafter.

Methods for generating chimeric and humanized antibodies also are known (see, e.g., U.S. patent No. 5,530,101 (Queen, et al.), U.S. patent No. 5,707,622 (Fung, et al.) and U.S. patent Nos.

5,994,524 and 6,245,894 (Matsushima, et al.)), which generally involve transplanting an antibody variable region from one species (e.g., mouse) into an antibody constant domain of another species (e.g., human). Antigen-binding regions of antibodies (e.g., Fab regions) include a light chain and a heavy chain, and the variable region is composed of regions from the light chain and the heavy chain. Given that the variable region of an antibody is formed from six complementarity-determining regions (CDRs) in the heavy and light chain variable regions, one or more CDRs from one antibody can be substituted (i.e., grafted) with a CDR of another antibody to generate chimeric antibodies. Also, humanized antibodies are generated by introducing amino acid substitutions that render the resulting antibody less immunogenic when administered to humans.

An antibody sometimes is an antibody fragment, such as a Fab, Fab', F(ab)'2, Dab, Fv or single-chain Fv (ScFv) fragment, and methods for generating antibody fragments are known (see, e.g., U.S. Patent Nos. 6,099,842 and 5,990,296 and PCT/GB00/04317). In some embodiments, a binding partner in one or more hybrids is a single-chain antibody fragment, which sometimes are constructed by joining a heavy chain variable region with a light chain variable region by a polypeptide linker (e.g., the linker is attached at the C-terminus or N-terminus of each chain) by recombinant molecular biology processes. Such fragments often exhibit specificities and affinities for an antigen similar to the original monoclonal antibodies. Bifunctional antibodies sometimes are constructed by engineering two different binding specificities into a single antibody chain and sometimes are constructed by joining two Fab' regions together, where each Fab' region is from a different antibody (e.g., U.S. Patent No. 6,342,221). Antibody fragments often comprise engineered regions such as CDR-grafted or humanized fragments. In certain embodiments the binding partner is an intact immunoglobulin, and in other embodiments the binding partner is a Fab monomer or a Fab dimer.

In some embodiments, one or more promoter elements preferentially active in the solid tumors of living organisms may be operably linked, on the same or different nucleic acid reagents, to nucleotide sequences that can encode one or more components of a multi-component (e.g., two or more components) therapeutic agent. Therapeutic agents for such applications include, without limitation, an enzyme coding sequence, a prodrug coding sequence; a protein comprising two peptide sequences that interact to form the therapeutic agent; related genes from a metabolic pathway; or one or more RNA molecules that functionally interact to form a therapeutic agent, for example. In certain embodiments targeted, tumor specific therapies may comprise an expression system that may comprise a nucleic acid reagent contained in a recombinant host cell. The term

"operably linked" as used herein refers to a nucleic acid sequence (e.g., a coding sequence) present on the same nucleic acid molecule as a promoter element and whose expression is under the control of said promoter element.

Expression Systems

Embodiments described herein provide an expression system useful for delivering a therapeutic agent or pharmaceutical composition (e.g., toxin, drug, prodrug, or microorganism (e.g. recombinant host cell) expressing a toxin, drug, or prodrug) to a specific target or tissue within a living subject exhibiting a condition treatable by the therapeutic agent or pharmaceutical composition (e.g., living organism with a solid tumor, for example). Embodiments described herein also may be useful for driving production of a system for generating toxic substances or to elicit responses from the host, for example by expressing cytokines, interleukins, growth inhibitors, or therapeutic RNA's or proteins from the expression system or causing the host organism to increase expression of cytokines, interleukins, growth inhibitors, or therapeutic RNA's or proteins by expression of an agent which can elicit the appropriate metabolic or immunological response. In some embodiments, the expression system may comprise a nucleic acid reagent and a delivery vector. The delivery vector sometimes can be a microorganism (e.g., bacteria, yeast, fungi, or virus) that harbors the nucleic acid reagent, and can express the product of the nucleic acid reagent or can deliver the nucleic acid reagent to the subject for expression within host cells.

In some embodiments, an expression system may comprise a promoter element operably linked to a therapeutic gene of a nucleic acid reagent. The nucleic acid reagent may be disposed in a bacterial host, where the bacterial host comprising the nucleic acid reagent is delivered to a eukaryotic organism such that expression of the nucleic acid reagent, in the appropriate tissue or structure (e.g., inside a solid tumor, for example) causes a therapeutic effect. In certain embodiments, the expression system promoter elements sometimes can be regulated (e.g., induced or repressed) in a eukaryotic environment (e.g., bacteria inside a eukaryotic organism or specific organ or structure in an organism). In some embodiments, the expression system promoter elements, isolated using methods described herein, can be selectively regulated. That is, the promoter elements sometimes can be influenced to increase transcription by providing the appropriate selective agent (e.g., administering tetracycline or kanomycin, metals, or starvation for a particular nutrient, for example, and described further below) to the host organism, such that the recombinant host cell containing the nucleic acid reagent comprising a selectable promoter

element responds by showing a demonstrable (e.g., at least two fold, for example) increase in transcription activity from the promoter element.

In certain embodiments, an expression system may comprise a nucleotide sequence encoding a toxic or therapeutic RNA or protein or an RNA or protein that participates in generating a toxin or therapeutic agent operably linked to a promoter identified by the methods described herein. In some embodiments, an expression system as described herein may comprise a first promoter nucleotide sequence operably linked to a first coding sequence and a second promoter nucleotide sequence operably linked to a second coding sequence, where: the first coding sequence and the second coding sequence may encode RNA or polypeptides that individually do not inhibit tumor growth; RNA or polypeptides encoded by the first coding sequence and the second coding sequence, in combination, inhibit tumor growth; and the first promoter nucleotide sequence and the second promoter nucleotide sequence can be preferentially activated in solid tumors of living organisms. In some embodiments an expression system as described herein may comprise two or more sequences encoding toxic or therapeutic RNA or proteins, or RNA or proteins that participate in generating a toxin or therapeutic agent, operably linked to a similar number of promoter elements identified by methods described herein.

In some embodiments, a nucleotide coding sequence can encode an RNA that has a function other than encoding a protein. Non-limiting examples of coding sequences that do not encode proteins are tRNA, rRNA, siRNA, or anti-sense RNA. rRNA's (e.g., ribosomal RNA's) of various organisms sometimes have point mutations that confer antibiotic resistance. Expression of rRNA's that contain antibiotic resistance mutations inside a solid tumor, when the rRNA's are operably linked to a heterologous promoter sequence isolated using methods described herein, may provide a method for ensuring the survival of the recombinant cells only in the tumor environment, due to the resistance phenotype induced in the solid tumors. Therefore, all recombinant cells carrying the expression system would be susceptible to the antibiotic administered to the organism, except in the inside of the solid tumor.

In some embodiments, there is provided an expression system described above, where the first coding sequence can encode an enzyme, the second coding sequence can encode a prodrug, and the enzyme can process the prodrug into a drug that inhibits tumor growth. A non-limiting example of this type of combination is an inactive peptide toxin and an enzyme which cleaves the inactive form to release the active form of the toxin. Another example may be an antibody, whose protein

sequence has been determined and a synthetic gene has been generated, and which requires processing (e.g., polypeptide cleavage) for assembly into an active form. In such examples, the first and second coding sequences are preferentially expressed inside the solid tumors, as the methods described herein select promoter elements preferentially activated in solid tumors. The combination of targeted, tumor specific expression, by delivery of the expression system comprising the nucleic acid reagent further comprising promoter elements preferentially activated in solid tumors of living organisms, as identified and isolated as described herein, and enzyme catalyzed activation of prodrugs, offers a significant improvement in gene-directed enzyme prodrug therapies. The expression systems described herein can be used to express prodrugs that, when activated, increase the bioavailability of therapeutic agents in solid tumor, or directly inhibit tumor growth by the action of the activated prodrug. In some embodiments, the second coding sequence can be a bacterial operon encoding a number of peptides, polypeptides or proteins which functionally form the prodrug. In some embodiments the first and second coding sequences can encode synthetically engineered enzymes or proteins specifically designed as prodrugs for anticancer therapies.

In some embodiments, there is provided an expression system, where the first coding sequence can encode a first polypeptide, the second coding sequence can encode a second polypeptide, and the first polypeptide and the second polypeptide form a complex that inhibits tumor growth. Non-limiting examples of two component protein or peptide toxins that can be used as therapeutic agents include Diphtheria toxin, various Pertussis toxins, Pseudomonas endotoxin, various Anthrax toxins, and bacterial toxins that act as superantigens (e.g., Staphylococcus aureus Exfoliatin B, for example). A combination of targeted, tumor specific expression, by delivery of an expression system comprising a nucleic acid reagent further comprising promoter elements preferentially activated in solid tumors as identified and isolated as described herein, and the use of two component protein or peptide toxins, offers a significant improvement in targeted, *in situ* delivery of anticancer therapies. Another example of a complex can include expressing two or more portions of an antibody (e.g., a light chain and a heavy chain), where the two or more portions can self assemble into a complex having antibody binding activity (e.g., antibody fragment).

In some embodiments, the promoter elements of the expression systems described herein (e.g., the first promoter nucleotide sequence, the second promoter nucleotide sequence, or both promoter nucleotide sequences) comprise (i) a nucleotide sequence of Table 2A, (ii) a functional promoter nucleotide sequence 80% or more identical to a nucleotide sequence of Table 2A, or (iii)

or a functional promoter subsequence of (i) or (ii). That is, a functional promoter nucleotide sequences that is at least 80% or more, 81% or more, 82% or more, 83% or more, 84% or more, 85% or more, 86% or more, 87% or more, 88% or more, 89% or more, 90% or more, 91% or more, 92% or more, 93% or more, 94% or more, 95% or more, 96% or more, 97% or more, 98% or more, or 99% or more identical to a nucleotide sequence of Table 2A. The term "identical" as used herein refers to two or more nucleotide sequences having substantially the same nucleotide sequence when compared to each other. One test for determining whether two nucleotide sequences or amino acids sequences are substantially identical is to determine the percent of identical nucleotide sequences or amino acid sequences shared.

Sequence identity can also be determined by hybridization assays conducted under stringent conditions. As use herein, the term "stringent conditions" refers to conditions for hybridization and washing. Stringent conditions are known to those skilled in the art and can be found in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. , 6.3.1-6.3.6 (1989). Aqueous and non-aqueous methods are described in that reference and either can be used. An example of stringent hybridization conditions is hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 50°C. Another example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 55°C. A further example of stringent hybridization conditions is hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 60°C. Often, stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 65°C. More often, stringency conditions are 0.5M sodium phosphate, 7% SDS at 65°C, followed by one or more washes at 0.2X SSC, 1% SDS at 65°C.

Calculations of sequence identity can be performed as follows. Sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). The length of a reference sequence aligned for comparison purposes is sometimes 30% or more, 40% or more, 50% or more, often 60% or more, and more often 70% or more, 80% or more, 90% or more, or 100% of the length of the reference sequence. The nucleotides or amino acids at corresponding nucleotide or polypeptide positions, respectively, are then compared among the two sequences. When a position in the first sequence

is occupied by the same nucleotide or amino acid as the corresponding position in the second sequence, the nucleotides or amino acids are deemed to be identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, introduced for optimal alignment of the two sequences. Comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. Percent identity between two amino acid or nucleotide sequences can be determined using the algorithm of Meyers & Miller, CABIOS 4: 11-17 (1989), which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4. Also, percent identity between two amino acid sequences can be determined using the Needleman & Wunsch, J. Mol. Biol. 48: 444-453 (1970) algorithm which has been incorporated into the GAP program in the GCG software package (available at the http address www.gcg.com), using either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. Percent identity between two nucleotide sequences can be determined using the GAP program in the GCG software package (available at http address www.gcg.com), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. A set of parameters often used is a Blossum 62 scoring matrix with a gap open penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

In some embodiments, the first promoter nucleotide sequence and the second nucleotide sequence can be in the same nucleic acid molecule (e.g., the same nucleic acid reagent, for example). In certain embodiments, the first promoter nucleotide sequence and the second nucleotide sequence can be in different nucleic acid molecule (e.g., different nucleic acid reagents, for example). In some embodiments, three or more promoters can be in the same nucleic acid molecule, and in certain embodiments, three or more promoters can be on different nucleic acid molecules. In some embodiments, an expression system may comprise functional promoter subsequences that are about 20 to about 150 nucleotides in length.

In some embodiments, the first promoter nucleotide sequence (e.g., promoter element) and the second promoter nucleotide sequence can be bacterial nucleotide sequences. In some embodiments, three or more promoter nucleotide sequences can be bacterial nucleotide sequences. In certain embodiments, the bacterial sequences are *Enterobacteriaceae* sequences, and in some embodiments, the *Enterobacteriaceae* sequences are *Salmonella* sequences. In some embodiments, the expression systems described herein are contained within recombinant

host cells. In certain embodiments, the cells can be *Enterobacteriaceae*. In some embodiments, the *Enterobacteriaceae* can be *Salmonella*, and in certain embodiments, the *Salmonella* can be avirulent *Salmonella*.

Nucleic Acids

A nucleic acid can comprise certain elements, which often are selected according to the intended use of the nucleic acid. Any of the following elements can be included in or excluded from a nucleic acid reagent. A nucleic acid reagent, for example, may include one or more or all of the following nucleotide elements: one or more promoter elements, one or more 5' untranslated regions (5'UTRs), one or more regions into which a target nucleotide sequence may be inserted (an "insertion element"), one or more target nucleotide sequences, one or more 3' untranslated regions (3'UTRs), and a selection element. A nucleic acid reagent can be provided with one or more of such elements and other elements (e.g., antibiotic resistance genes, multiple cloning sites, and the like) can be inserted into the nucleic acid reagent before the nucleic acid is introduced into a suitable expression host or system (e.g., *in vivo* expression in host, or in vitro expression in a cell free expression system, for example). The elements can be arranged in any order suitable for expression in the chosen expression system.

In some embodiments, a nucleic acid reagent may comprise a promoter element where the promoter element comprises two distinct transcription initiation start sites (e.g., two promoters within a promoter element, for example). In some embodiments, a promoter element in a nucleic acid reagent may comprise two promoters. In certain embodiments, the promoter element may comprise a constitutive promoter and an inducible promoter, and in some embodiments a promoter element may comprise two inducible promoters. In certain embodiments a nucleic acid reagent may comprise two or more distinct or different promoter elements. In some embodiments, the promoters may respond to the same or different inducers or repressors of transcription (e.g., induce or repress expression of a nucleic acid reagent from the promoter element). A nucleic acid reagent sometimes can contain more than one promoter element that is turned on at specific times or under specific conditions.

A nucleic acid reagent sometimes can comprise a 5' UTR that may further comprise one or more elements endogenous to the nucleotide sequence from which it originates, and sometimes includes one or more exogenous elements. A 5' UTR can originate from any suitable nucleic acid, such as genomic DNA, plasmid DNA, RNA or mRNA, for example, from any suitable organism

(e.g., virus, bacterium, yeast, fungi, plant, insect or mammal). The artisan may select appropriate elements for the 5' UTR based upon the expression system being utilized. A 5' UTR sometimes comprises one or more of the following elements known to the artisan: enhancer sequences, silencer sequences, transcription factor binding sites, accessory protein binding site, feedback regulation agent binding sites, Pribnow box, TATA box, -35 element, E-box (helix-loop-helix binding element), transcription initiation sites, translation initiation sites, ribosome binding site and the like. In some embodiments, a promoter element may be isolated such that all 5' UTR elements necessary for proper conditional regulation are contained in the promoter element fragment, or within a functional sub sequence of a promoter element fragment.

A nucleic acid reagent sometimes can have a 3' UTR that may comprise one or more elements endogenous to the nucleotide sequence from which it originates, and sometimes includes one or more exogenous elements. A 3' UTR can originate from any suitable nucleic acid, such as genomic DNA, plasmid DNA, RNA or mRNA, for example, from any suitable organism (e.g., virus, bacterium, yeast, fungi, plant, insect or mammal). The artisan may select appropriate elements for the 3' UTR based upon the expression system being utilized. A 3' UTR sometimes comprises one or more of the following elements, known to the artisan, which may influence expression from promoter elements within a nucleic acid reagent: transcription regulation site, transcription initiation site, transcription termination site, transcription factor binding site, translation regulation site, translation termination site, translation initiation site, translation factor binding site, ribosome binding site, replicon, enhancer element, silencer element and polyadenosine tail. A 3' UTR sometimes includes a polyadenosine tail and sometimes does not, and if a polyadenosine tail is present, one or more adenosine moieties may be added or deleted from it (e.g., about 5, about 10, about 15, about 20, about 25, about 30, about 35, about 40, about 45 or about 50 adenosine moieties may be added or subtracted).

A nucleic acid reagent that is part of an expression system sometimes comprises a nucleotide sequence adjacent to the nucleic acid sequence encoding a therapeutic agent or pharmaceutical composition that is translated in conjunction with the ORF and encodes an amino acid tag. The tag-encoding nucleotide sequence is located 3' and/or 5' of an ORF in the nucleic acid reagent, thereby encoding a tag at the C-terminus or N-terminus of the protein or peptide encoded by the ORF. Any tag that does not abrogate transcription and/or translation may be utilized and may be appropriately selected by the artisan.

A tag sometimes comprises a sequence that localizes a translated protein or peptide to a component in a system, which is referred to as a "signal sequence" or "localization signal sequence" herein. A signal sequence often is incorporated at the N-terminus of a target protein or target peptide, and sometimes is incorporated at the C-terminus. Examples of signal sequences are known to the artisan, are readily incorporated into a nucleic acid reagent, and often are selected according to the expression chosen by the artisan. A tag sometimes is directly adjacent to an amino acid sequence encoded by a nucleic acid reagent (i.e., there is no intervening sequence) and sometimes a tag is substantially adjacent to the amino acid sequence encoded by the nucleic acid reagent (e.g., an intervening sequence is present). An intervening sequence sometimes includes a recognition site for a protease, which is useful for cleaving a tag from a target protein or peptide. A signal sequence or tag, in some embodiments, localizes a translated protein or peptide to a cell membrane.

Examples of signal sequences include, but are not limited to, a nucleus targeting signal (e.g., steroid receptor sequence and N-terminal sequence of SV40 virus large T antigen); mitochondria targeting signal (e.g., amino acid sequence that forms an amphipathic helix); peroxisome targeting signal (e.g., C-terminal sequence in YFG from *S.cerevisiae*); and a secretion signal (e.g., N-terminal sequences from invertase, mating factor alpha, PHO5 and SUC2 in *S.cerevisiae*; multiple N-terminal sequences of *B. subtilis* proteins (e.g., Tjalsma et al., Microbiol.Molec. Biol. Rev. 64: 515-547 (2000)); alpha amylase signal sequence (e.g., U.S. Patent No. 6,288,302); pectate lyase signal sequence (e.g., U.S. Patent No. 5,846,818); precollagen signal sequence (e.g., U.S. Patent No. 5,712,114); OmpA signal sequence (e.g., U.S. Patent No. 5,470,719); lam beta signal sequence (e.g., U.S. Patent No. 5,389,529); *B. brevis* signal sequence (e.g., U.S. Patent No. 5,232,841); and *P. pastoris* signal sequence (e.g., U.S. Patent No. 5,268,273)).

A nucleic acid reagent sometimes contains one or more origin of replication (ORI) elements. In some embodiments, a template comprises two or more ORIs, where one functions efficiently in one organism (e.g., a bacterium) and another functions efficiently in another organism (e.g., a eukaryote). A nucleic acid reagent often includes one or more selection elements. Selection elements often are utilized using known processes to determine whether a nucleic acid reagent is included in a cell. In some embodiments, a nucleic acid reagent includes two or more selection elements, where one functions efficiently in one organism and another functions efficiently in another organism.

Examples of selection elements include, but are not limited to, (1) nucleic acid segments that encode products that provide resistance against otherwise toxic compounds (e.g., antibiotics); (2) nucleic acid segments that encode products that are otherwise lacking in the recipient cell (e.g., essential products, tRNA genes, auxotrophic markers); (3) nucleic acid segments that encode products that suppress the activity of a gene product; (4) nucleic acid segments that encode products that can be readily identified (e.g., phenotypic markers such as antibiotics (e.g., β-lactamase), β-galactosidase, green fluorescent protein (GFP), yellow fluorescent protein (YFP), red fluorescent protein (RFP), cyan fluorescent protein (CFP), and cell surface proteins); (5) nucleic acid segments that bind products that are otherwise detrimental to cell survival and/or function; (6) nucleic acid segments that otherwise inhibit the activity of any of the nucleic acid segments described in Nos. 1-5 above (e.g., antisense oligonucleotides); (7) nucleic acid segments that bind products that modify a substrate (e.g., restriction endonucleases); (8) nucleic acid segments that can be used to isolate or identify a desired molecule (e.g., specific protein binding sites); (9) nucleic acid segments that encode a specific nucleotide sequence that can be otherwise non-functional (e.g., for PCR amplification of subpopulations of molecules); (10) nucleic acid segments that, when absent, directly or indirectly confer resistance or sensitivity to particular compounds; (11) nucleic acid segments that encode products that either are toxic (e.g., Diphtheria toxin) or convert a relatively non-toxic compound to a toxic compound (e.g., Herpes simplex thymidine kinase, cytosine deaminase) in recipient cells; (12) nucleic acid segments that inhibit replication, partition or heritability of nucleic acid molecules that contain them; and/or (13) nucleic acid segments that encode conditional replication functions, e.g., replication in certain hosts or host cell strains or under certain environmental conditions (e.g., temperature, nutritional conditions, and the like).

Nucleic acid reagents can comprise naturally occurring sequences, synthetic sequences, or combinations thereof. Certain nucleotide sequences sometimes are added to, modified or removed from one or more of the nucleic acid reagent elements, such as the promoter, 5'UTR, target sequence, or 3'UTR elements, to enhance or potentially enhance transcription and/or translation before or after such elements are incorporated in a nucleic acid reagent. Certain embodiments are directed to a process comprising: determining whether any nucleotide sequences that increase or potentially increase transcription efficiency are not present in the elements, and incorporating such sequences into the nucleic acid reagent. A nucleic acid reagent can be of any form useful for the chosen expression system.

In some embodiments, a nucleic acid reagent sometimes can be an isolated nucleic acid molecule which may comprise a recombinant expression system, which expression system can comprise a nucleotide sequence encoding a toxic or therapeutic RNA or protein, or an RNA or protein that participates in generating a toxin or therapeutic agent operably linked to a heterologous promoter which promoter is preferentially activated in solid tumors in living organisms. In some embodiments, the promoter sequence can be a naturally occurring nucleotide sequence. In certain embodiments, a nucleic acid reagent sometimes can be two or more isolated nucleic acid molecules which may comprise a recombinant expression system, which expression system can comprise two or more nucleotide sequences encoding toxic or therapeutic RNA's or proteins, or RNA's or proteins that participate in generating a toxin or therapeutic agent operably linked to two or more heterologous promoters which promoters is preferentially activated in solid tumors in living organisms. In some embodiments, the isolated nucleic acid of the recombinant expression system is a promoter nucleic acid. In certain embodiments, the promoter is an *Enterobacteriaceae* promoter, and in some embodiments, the promoter is a *Salmonella* promoter.

Promoters

A promoter element typically comprises a region of DNA that can facilitate the transcription of a particular gene, by providing a start site for the synthesis of RNA corresponding to a gene. Promoters often are located near the genes they regulate, are located upstream of the gene (e.g., 5' of the gene), and are on the same strand of DNA as the sense strand of the gene, in some embodiments. A promoter often interacts with a RNA polymerase, an enzyme that catalyses synthesis of nucleic acids using a preexisting nucleic acid. When the template is a DNA template, an RNA molecule is transcribed before protein is synthesized. Promoter elements can be found in prokaryotic and eukaryotic organisms

A promoter element generally is a component in an expression system comprising a nucleic acid reagent. An expression system often can comprise a nucleic acid reagent and a suitable host for expression of the nucleic acid reagent. For example, an expression system may comprise a heterologous promoter operably linked to a toxin gene, carried on a nucleic acid reagent that is expressed in a bacterial host, in some embodiments. Promoter elements isolated using methods described herein may be recognized by any polymerase enzyme, and also may be used to control the production of RNA of the therapeutic agent or pharmaceutical composition operably linked to the promoter element in the nucleic acid reagent. In some embodiments, additional 5' and/or 3'

UTR's may be included in the nucleic acid reagent to enhance the efficiency of the isolated promoter element.

Methods described herein can be used to identify a promoter preferentially activated in tumor tissue. In some embodiments the method comprises; (a) providing a library of expression systems each comprising a nucleotide sequence encoding a detectable protein operably linked to a different candidate promoter; (b) providing the library to solid tumor tissue and to normal tissue; (c) identifying cells from each tissue that show high levels of expression of the detectable protein; and (d) obtaining the expression systems from the cells that produce greater levels of detectable protein in tumor tissue as compared to normal tissue, and identifying the promoters of the expression system. In some embodiments, the method further comprises scoring the promoters identified in (d) (e.g., by detecting a detectable protein, GFP for example). In certain embodiments, the library is provided in recombinant host cells. In some embodiments, the library of DNA fragments ranged in size from about 25 base pairs to about 10,000 base pairs in length. In some embodiments, the fragments can be randomly sized fragments. In certain embodiments, the fragments can be an ordered set of specific sequences in a particular size range.

In some embodiments, the promoters are *Salmonella* promoters and the recombinant host cells are *Salmonella*. In certain embodiments, the candidate promoters are from bacteria, or are 80% or more identical to promoters from bacteria. That is, the candidate promoters can be at least 80% or more, 81% or more, 82% or more, 83% or more, 84% or more, 85% or more, 86% or more, 87% or more, 88% or more, 89% or more, 90% or more, 91% or more, 92% or more, 93% or more, 94% or more, 95% or more, 96% or more, 97% or more, 98% or more, or 99% or more identical to promoters from bacteria. In some embodiments, the bacteria are *Enterobacteriaceae* (e.g., *Salmonella*).

Detailed experimental procedures for construction of promoter trap constructs and libraries are presented below in Example 1 and in FIG. 1. FIG. 1 is a flow diagram outlining how the libraries were enriched for promoter sequences preferentially activated in solid tumors. The initial library was constructed by ligating sonicated, end repaired *Salmonella* genomic DNA, size selected for fragments 300 to 500 base pairs in length into a promoter trap construct upstream of a promoterless green fluorescent protein (GFP) sequence. Although GFP was the detectable protein used herein, due to ease of detection, any detectable protein that can be easily and efficiently detected can be used in place of GFP. Non-limiting examples of detectable proteins are other

fluorescent proteins, peptides or proteins that inactivate antibiotics (e.g., beta-lactamase, the enzyme responsible for penicillin resistance, for example) and the like.
The library contained in recombinant cells can be injected into rodents (e.g., mice, rats) bearing solid tumor xenografts, as described below. Enrichment for promoters preferentially active in tumors was performed as described in Example 2. The experimental results from the enrichment process are presented in Tables 2-7. Tables 2-7 contain sequences of promoters active in normal tissue (e.g., spleen), promoters active in both normal tissue and solid tumors and promoters preferentially activated in solid tumors (see Tables 2A, 2B, 6A and 6B).

The sequences isolated using the methods described herein were mapped to genome positions as described in Example 2, using high density, high resolution arrays constructed as described in Example 1. The nucleotide position of the library construct that had the highest enrichment signal for a particular library construct is given in the Tables as the nucleotide position. The nucleotide position may correspond to the start site of the isolated promoter element. Definitive promoter start site mapping can be performed using a suitable method. One method is 5' RACE (e.g., rapid amplification of cDNA ends), for example, which can be routinely performed. 5' RACE can be used to identify the first nucleotide in an mRNA or other RNA molecule and also be used to identify and/or clone a gene when only a small portion of the sequence is known. An example of a 5' RACE procedure suitable for identifying a transcription start site from promoter elements isolated using the methods described herein is Schramm et al, "A simple and reliable 5' RACE approach", Nucleic Acids Research, 28(22):e96, 2000.

Where identifiable, gene names and functions are presented along with the sequence information for the isolated nucleic acid sequences that exhibited promoter activity (e.g., showed at least a two fold increase in detectable GFP over input). Table 6 describes the distribution of sequences isolated using the methods described herein. The majority of sequences that exhibited promoter activity (e.g., transcription of GFP) were isolated from intergenic sequences. This observation is in keeping with the finding that many bacterial promoters lie outside of gene coding sequences. Further distribution results are discussed in Example 2.

To confirm the tumor specificity of the isolated sequences, a number of clones were further investigated (see Example 2, Confirmation of tumor specificity *in vivo*). In particular, Clone ID Nos. 10, 28, 45, 44, and 84 were further investigated *in vivo* as described in Example 2. Three clones in particular were induced to a greater degree in tumor as compared to spleen (e.g., Clones 10, 28

and 45). FIG. 2 illustrates the expression of GFP from these clones *in vivo* in whole mice and in tumor alone. FIG. 2 presents the microscopic imaging (Olympus OV100 small animal imaging system) of fluorescent bacteria in mouse spleen and tumors. Clone C28 maps to the upstream intergenic region of the *flhB* gene, clone C10 maps to the *pefL* intergenic region, and C45 maps to the intergenic region of the gene *ansB*. The number of colony forming units for each trial is given below the image, to account for differences in signal intensities. The number of colony forming units isolated in each trial was approximately equal, and therefore did not contribute to the differences in intensity seen in the images.

Certain promoter elements can be regulated in a conditional manner. That is, promoters sometimes can be turned on, turned off, up-regulated or down-regulated by the influence of certain environmental, nutritional, or internal signals (e.g., heat inducible promoters, light regulated promoters, feedback regulated promoters, hormone influenced promoters, tissue specific promoters, oxygen and pH influenced promoters and the like, for example). Promoters influenced by environmental, nutritional or internal signals frequently are influenced by a signal (direct or indirect) that binds at or near the promoter and increases or decreases expression of the target sequence under certain conditions and/or in specific tissues. Certain promoter elements can be regulated in a selective manner, as noted above. In some embodiments, the promoter does not include a nucleotide sequence to which a bacterial (e.g., gram negative (e.g., E. coli, Salmonella) oxygen-responsive global transcription factor (FNR) binds substantially. In certain embodiments, the promoter sequence does not include one or more of the following subsequences: GGATAAAAGTGACCTGACGCAATATTTGTCTTTTCTTGCTTAATAATGTTGTCA, GGATAAAAGTGACCTGACGCAATATTTGTCTTTTCTTGCTTTATAATGTTGTCA, GGATAAAATTGATCTGAATCAATATTTGTCTTTTCTTGCTTAATAATGTTGTCA, or GGATAAAAGGATCCGACGCAATATTGTCTTTTCTTGCTTAATAATGTTGTCA. In some embodiments, the promoter sequence is not identical to a bacterial promoter that regulates the bacterial *pepT* gene.

Non-limiting examples of selective agents that can be used to selectively regulate promoters in therapeutic methods using expression systems and promoter elements described herein include, (1) nucleic acid segments that encode products that provide resistance against otherwise toxic compounds (e.g., antibiotics); (2) nucleic acid segments that encode products that are otherwise lacking in the recipient cell (e.g., essential products, tRNA genes, auxotrophic markers); (3) nucleic acid segments that encode products that suppress the activity of a gene product; (4) nucleic acid

segments that encode products that can be readily identified (e.g., phenotypic markers such as antibiotics (e.g., β-lactamase), β-galactosidase, green fluorescent protein (GFP), yellow fluorescent protein (YFP), red fluorescent protein (RFP), cyan fluorescent protein (CFP), and cell surface proteins); (5) nucleic acid segments that bind products that are otherwise detrimental to cell survival and/or function; (6) nucleic acid segments that otherwise inhibit the activity of any of the nucleic acid segments described in Nos. 1-5 above (e.g., antisense oligonucleotides); (7) nucleic acid segments that bind products that modify a substrate (e.g., restriction endonucleases); (8) nucleic acid segments that can be used to isolate or identify a desired molecule (e.g., specific protein binding sites); (9) nucleic acid segments that encode a specific nucleotide sequence that can be otherwise non-functional (e.g., for PCR amplification of subpopulations of molecules); (10) nucleic acid segments that, when absent, directly or indirectly confer resistance or sensitivity to particular compounds; (11) nucleic acid segments that encode products that either are toxic (e.g., Diphtheria toxin) or convert a relatively non-toxic compound to a toxic compound (e.g., Herpes simplex thymidine kinase, cytosine deaminase) in recipient cells; (12) nucleic acid segments that inhibit replication, partition or heritability of nucleic acid molecules that contain them; and/or (13) nucleic acid segments that encode conditional replication functions, e.g., replication in certain hosts or host cell strains or under certain environmental conditions (e.g., temperature, nutritional conditions, and the like). In some embodiments, the nucleic acids identified and isolated using methods described herein (e.g., promoter elements preferentially activated in solid tumors of living organisms) can be selectively regulated by administration of a suitable selective agent, as described above or known and available to the artisan.

Methods presented herein take into account the unique environment inside a tumor. Therefore, while hypoxia induced tumors may be identified, other promoters preferentially activated in the unique tumor environment can also be identified and isolated. Some specific classes of promoters preferentially activated inside tumors were presented above. Therefore, the promoters isolated using methods described herein may be preferentially activated under a wide variety of regulatory molecules and conditions.

Therapeutic Agents and Methods of Treatment

Expression systems, nucleic acid reagents and pharmaceutical compositions described herein that comprise promoter elements preferentially activated in solid tumors, or cells containing the expression system, nucleic acid reagents and pharmaceutical compositions described herein, can

be used to treat solid tumors in a living organism. In some embodiments, methods for treating solid tumors comprise administering to a subject harboring the tumors the nucleic acid molecules or nucleic acid reagents comprising nucleic acid sequences preferentially activated in tumors (e.g., nucleic acids bearing promoter elements isolated using the methods described herein, for example), cells containing the above described nucleic acids, or compositions comprising the isolated nucleic acids. In some embodiments, the expression system, nucleic acid reagent, and/or pharmaceutical compositions comprise a nucleotide sequence encoding a toxic or therapeutic RNA or protein, or an RNA or protein that participates in generating a desired toxin or therapeutic agent operably linked to a promoter identified by the methods described herein.

In some embodiments, the therapeutic RNA or protein can be an enzyme which catalyzes the activation of a prodrug. That is, the enzyme can be operably linked to a promoter element preferentially activated in solid tumors. The nucleic acid reagent / expression system / pharmaceutical composition contained in a recombinant cell can be administered along with the prodrug (e.g., administered by intramuscular or intravenous injection, for example). The avirulent recombinant host cell sometimes can preferentially colonize the solid tumor, and the prodrug will remain inactive in all tissues except inside the solid tumor, due to the enzyme only being produced by recombinant cells that have colonized the tumor, due to the heterologous promoter that is preferentially activated in the solid tumors of living organisms. Non-limiting examples of this type of combination are the enzymes nitroreductase or quinone reductase 2 and the prodrug CB1954 (5-[aziridin-1-yl]-2,4-dinitrobenzamide), or Cytochrome P450 enzymes 2B1, 2B4, and 2B5 and the anticancer prodrugs Cyclphosphamide and Ifosfamide. Further non-limiting examples of enzyme prodrug combinations can be found in Rooseboom et al, "Enzyme-Catalyzed Activation of Anticancer Prodrugs", Pharmacol. Rev. 56:53-102, 2004, hereby incorporated by reference in its entirety.

In certain embodiments, bacterial two component toxins can also be utilized as the toxic or therapeutic proteins or peptide sequences operably linked to the promoters isolated using methods described herein. Non-limiting examples of bacterial toxins suitable for use in compositions described herein were presented above. Several of these toxins offer attractive modes of toxicity that when combined with the expression only inside a solid tumor, may offer novel therapies for inhibiting tumor growth. For example, Diphtheria toxin and Pseudomonas Exotoxin A are both two component toxins (e.g., has two distinct peptides) that inhibit protein synthesis, resulting in cell death. The nucleic acid sequences of these toxins could be operably linked to promoters

preferentially activated in solid tumors, and administered to a subject harboring a solid tumor, with little or no toxicity to the organism outside of the targeted solid tumor.

In some embodiments, multiple nucleic acid reagents can be administered, where each nucleic acid reagent comprises a nucleic acid sequence for a gene in a metabolic pathway, the pathway producing a therapeutic agent that can inhibit tumor growth. In certain embodiment the nucleic acid reagents can have the same or different heterologous promoters preferentially activated in tumors operably linked to the sequences for the metabolic pathway genes.

In certain embodiments, the expression systems described herein may generate RNA's or proteins that are themselves toxic, or RNA's or proteins that are known to have a therapeutic effect by selective toxicity to solid tumors. A non-limiting example of a protein known to have a therapeutic effect by selective toxicity to solid tumors is Methioninase, which is known to be selectively inhibitory to tumors. Additional known toxic proteins include, but are not limited to, ricin, abrin, and the like. In addition to proteins that are toxic per se, the expression systems may generate proteins that convert non-toxic compounds into toxic ones. A non-limiting example is the use of lyases to liberate selenium from selenide analogs of sulfur-containing amino acids. Other non-limiting examples include generation of enzymes that liberate active compounds from inactive prodrugs. For example, derivatized forms of palytoxin can be provided that are non-toxic and the expression system used to produce enzymes that convert the inactive form to the toxic compound. In addition, proteins that attract systems in the host can also be expressed, including immunomodulatory proteins such as interleukins.

The subjects that can benefit from the embodiments, methods and compositions described herein include any subject that harbors a solid tumor in which the promoter operably linked to a therapeutic agent is preferentially active. Human subjects can be appropriate subjects for administering the compositions described herein. The methods and compositions described herein can also be applied to veterinary uses, including livestock such as cows, pigs, sheep, horses, chickens, ducks and the like. The methods and compositions described herein can also be applied to companion animals such as dogs and cats, and to laboratory animals such as rabbits, rats, guinea pigs, and mice.

The tumors to be treated include all forms of solid tumor, including tumors of the breast, ovary, uterus, prostate, colon, lung, brain, tongue, kidney and the like. Localized forms of highly metastatic tumors such as melanoma can also be treated in this manner.

Thus, the methods and compositions described herein may provide a selective means for producing a therapeutic or cytotoxic effect locally in tumor or other target tissue. As the encoded RNA's or proteins are produced uniquely or preferentially in tumor tissue, side effects due to expression in normal tissue is minimized.

Nucleic acid molecules may be formulated into pharmaceutical compositions for administration to subjects. The nucleic acid molecules sometimes are transfected into suitable cells that provide activating factors for the promoter. In some cases, the tumor cells themselves may contain workable activators. If the promoter is a bacterial promoter, bacteria, such as Salmonella itself, may be used. Any cell closely related to that from which the promoter derives is a suitable candidate. A preferred mode of administration is the use of bacteria that preferentially reside in hypoxic environments of solid tumors. The compositions which contain the nucleic acids, vectors, bacteria, cells, etc., sometimes are administered parenterally, such as through intramuscular or intravenous injection. The compositions can also be directly injected into the solid tumor. Nucleic acids sometimes are administered in naked form or formulated with a carrier, such as a liposome. A therapeutic formulation may be administered in any convenient manner, such as by electroporation, injection, use of a gene gun, use of particles (e.g., gold) and an electromotive force, or transfection, for example. Compositions may be administered in vivo, ex vivo or in vitro, in certain embodiments.

As noted above, ancillary substances may also be needed such as compounds which activate inducible promoters, substrates on which the encoded protein will act, standard drug compositions that may complement the activity generated by the expression systems of the invention and the like. These ancillary components may be administered in the same composition as that which contains the expression system or as a separate composition. Administration may be simultaneous or sequential and may be by the same or different route. Some ancillary agents may be administered orally or through transdermal or transmucosal administration.

The pharmaceutical compositions may contain additional excipients and carriers as is known in the art. Suitable diluents and carriers are found, for example, in *Remington's Pharmaceutical Sciences*, latest edition, Mack Publishing Co., Easton, PA, incorporated herein by reference.

Examples

The examples set forth below illustrate certain embodiments and do not limit the invention.

Example 1: Materials and Methods

Vector Construction.

Promoter trap plasmids with TurboGFP (e.g., promoter reporter plasmid comprising a destabilized TurboGFP, World Wide Web URL evrogen.com/TurboGFP.shtml) were generated by PCR from the pTurboGFP plasmid. The pTurboGFP plasmid was PCR amplified using the primers Turbo-LVA R1 (SEQ ID NO. 1, see Table 1) and Turbo-F1 (SEQ ID NO. 2, see Table 1) to generate a fusion of the peptide motif AANDENYALVA (SEQ ID NO. 3) to the 3' end of the protein (Andersen et al., 1998; Keiler and Sauer, 1996). The PCR product was digested by EcorRV and self ligated to generate pTurboGFP- LVA. The plasmids pTurboGFP and pTurboGFP-LVA were each double digested by XhoI and BamH1 to remove the T5 promoter sequence. The pairs of oligos PR1-1F / PR1-1R (SEQ ID NOS. 4 and 5, respectively, see Table 1) and PRL3-1F / PR3-1R (SEQ ID NOS. 6 and 7, respectively, see Table 1), containing multi-cloning sites, transcriptional terminators, and a ribosomal binding site, were used to replace the T5 constitutive promoter of pTurbo-GFP and pTurboGFP-LVA respectively. Primers Turbo-4F and Turbo-1R (SEQ ID NOS. 8 and 9, respectively, see Table 1) were used to amplify promoter inserts before and after FACS sort.

Table 1. Sequences of oligonucleotides use to construct promoter trap constructs

Oligos	Sequence
Turbo-LVA R1	SEQ.ID.NO. 1: ACTGATATCTTAAGCTACTAAAGCGTAGTTTTCGTCGTTTGCTGCAGGCCTT TCTTCACCGGCATCTGCA
T urbo-F1	SEQ.ID.NO. 2: CTGATATCGCTTGGACTCCTGTTGATAGAT
PRL1-1F	SEQ.ID.NO. 4: TCGAGAGATCTCCATCGAATTCGTGGGTCGACCCCGGGAGGCCTAAAGAG GAGAAATTAACTATGAGAGGATCGG
PRL1-1R	SEQ.ID.NO. 5: GATCCCGATCCTCTCATAGTTAATTTCTCCTCTTTAGGCCTCCCGGGGTCGA CCCACGAATTCGATGGAGATCTC
PRL3-1F	SEQ.ID.NO. 6: TCGAGCGAAATTAATACGACTCACTATAGGGAGACCCCCGGGTTAACACTA GTAAAGAGGAGAAATTAACTATGAGAGGATCGG
PRL3-1R	SEQ.ID.NO. 7: GATCCCGATCCTCTCATAGTTAATTTCTCCTCTTTACTAGTGTTAACCCGGG GGTCTCCCTATAGTGAGTCGTATTAATTTCGC
Turbo-4F	SEQ.ID.NO. 8: AAAGTGCCACCTGACGTCT
Turbo-1R	SEQ.ID.NO. 9: CCACCAGCTCGAACTCCAC

Promoter Library Construction.

10 µg of *Salmonella enterica* serovar *typhimurium* 14028 (*S. enterica. Typhimurium* 14028, ATCC) genomic DNA was eluted in TE buffer and sonicated with 3 pulses for 5 seconds on ice. Sonicated DNA was precipitated with 2 volumes ethanol and 0.1 volumes of Sodium Acetate (100 mM) and separated on a 1% agarose gel. 300 to 500 base pair (bp) fragments were recovered from the gel and DNA ends were repaired by T4 DNA polymerase. Repaired fragments were cloned in a dephosphorylated promoterless GFP plasmid upstream of a StuI and HpaI restriction site in the stable and destabilized GFP, respectively. These fragments were located just upstream of the GFP start codon, and were therefore capable of promoting transcription, depending on their sequence properties. The number of independent clones was approximately 120,000 for the stable variant and 60,000 for the unstable variant. The two libraries were mixed 1:1 and

designated "Library-0". This library contained about 180,000 independent Typhimurium fragments, representing about 15-fold coverage of the 4.8 Mb genome with clone spacing averaging every 25 bases. Hybridization to a *Salmonella* array showed that library-0 represented sequences from almost the entire genome.

Array Design.

A high-resolution array was generated using Roche NimbleGen high definition array technology (World Wide Web URL nimblegen.com/products/index.html). The array comprised 387,000 46-mer to 50-mer oligonucleotides, with length adjusted to generate similar predicted melting temperatures (Tm). 377,230 of these probes were designed based on the Typhimurium LT2 genome (NC-003197; McClelland et al, "Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2", Nature 413:852-856, 2001). Oligonucleotides tiled the genome every 12 bases, on alternating strands. Thus, each base pair in the genome was represented in four to six oligonucleotides, with two to three oligonucleotides on each strand. Probes representing the three LT2 regions not present in the genome of the very closely related 14028s strain (phages Fels-1 and Fels-2, STM3255-3260) and greater than 9,000 other oligonucleotides were included as controls for hybridization performance, synthesis performance, and grid alignment. The oligonucleotides were distributed in random positions across the array.

Fluorescence Activated Cell Sorting (FACS) Analysis.

Bacteria harboring the constitutive pTurboGFP plasmid were used as a positive control for the Becton Dickinson FACSAria FACS system. Side scatter ssc-w (*X*-axis) and ssc-H (*Y*-axis) were used to gate on single bacterial cells. GFP-fluorescence (GFP-A) on the *X*-axis and auto-fluorescence (PE) on the *Y*-axis permitted discrimination between green *Salmonella* cells and other fluorescent particles of different sizes. Fluorescent particles tended to be distributed on the diagonal of the GFP-A/PE plot, and had a fluorescence/auto-fluorescence ratio close to 1. Individual GFP-positive *Salmonella* cells had a higher ratio of fluorescence/auto-fluorescence and tended to be distributed close to the X-axis of the GFP-A/PE plot. Putative GFP-positive events in the window enriched for GFP-expressing *Salmonella* were sorted at a speed of `5,000 total events per second.

Example 2: Experimental Results

Enrichment of Active Promoters in Spleen.

To identify active *Salmonella* promoters in the spleen, five tumor-free nude mice were i.v. injected with 10^7 colony forming units (cfu) of *Salmonella* carrying a promoter library. This library, designated "library-0", consisted of ~180,000 plasmid clones each containing a fragment of the *Salmonella* genome upstream of a promoterless GFP gene (described above). Two days after injection, spleens were combined, homogenized on ice, and treated thrice with PBS containing 0.1% Triton X -100. An aliquot of the final homogenized sample was plated on Luria-Bertani (LB) medium with 50 μg/mL of ampicillin (Amp) to determine the number of bacterial colony-forming units (cfu). The remainder of the bacteria in the sample was immediately separated by FACS. Fifty thousand potentially GFP-positive events were sorted and this sublibrary was grown overnight in LB+ Amp and designated "library-1". The spleen was chosen because it is the primary site of *Salmonella* accumulation in normal mice (Ohl and Miller, "*Salmonella*: a model for bacterial pathogenesis", Annu. Rev. Med. 52:259-274, 2001).

Enrichment of Active Promoters in Tumor.

The experimental design for tumor samples is described in FIG. 1. Five nude mice bearing human-PC3 prostate tumors, between 0.5 and 1 cm^3 in size, were injected intratumorally with 10^7 cfu of *Salmonella* promoter library-0. Two days after injection, tumors were combined, homogenized on ice and washed, as above. An aliquot was plated to determine the number of bacterial colony-forming units. The remainder of the sample was immediately separated by FACS. Fifty thousand GFP-positive events were recovered and grown overnight in LB containing ampicillin (library-2). A small aliquot of these bacteria were then pelleted and resuspended in PBS (10^6 cfu/mL) and FACS sorted. GFP-negative events (10^6) were collected, grown in LB overnight, washed in PBS and reinjected into five human-PC3 tumors in nude mice. After 2 days, bacteria were extracted from tumors and 50,000 GFP-positive events were FACS sorted and expanded in LB+ Amp (library-3). A biological replicate of library-3 was obtained by repeating the experiment from the beginning using library-0. This was designated library-4.

Genome wide Survey on Tumor-Activated Promoters Using Arrays.

Plasmid DNA was extracted from the original promoter library (library-0), from clones activated in spleen (library-1), and from clones activated in subcutaneous PC3 tumors in nude mice after one (library-2) or two passages (library-3 and library-4) in tumors. Promoter sequences were recovered by PCR using primers Turbo-4F and Turbo-1R (see Table 1, presented above), and the PCR product was labeled by CY 5 (library-0) and CY 3 (library-1, library-2, library-3, library-4). The resulting products were then hybridized to the array of 387,000 oligonucleotide sequences (described above in Array Design) positioned at 12-base intervals around the Typhimurium genome (using the manufacturer's protocol) (Panthel et al, "Prophylactic anti-tumor immunity against a murine fibrosarcoma triggered by the *Salmonella* type III secretion system", Microbes Infect. 8:2539-2546, 2006). Spot intensities were normalized based on total signal in each channel. The enrichment of genomic regions was measured by the intensity ratio of the tumor or the spleen sample versus the input library (library-0). A moving median of the ratio of tumor versus input library from 10 data points (~170 bases) was calculated across the genome.

The highest median of each intergenic and intragenic region was chosen to represent the most highly overrepresented region of that promoter or gene in the tested library. Using a threshold of (exp / control) greater than or equal to 2, and enrichment in both replicates of the experiment (library-4, plus at least one of library-2 or library-3), there were 86 intergenic regions enriched in tumors but not in the spleen (see Table 2A and 2B, presented below), and 154 intergenic regions enriched in both tumor and spleen (see Table 3A and 3B, presented below). There were at least 30 regions enriched in spleen alone (see Table 4, presented below).

Table 2A. Intergenic regions that induce higher GFP expression in tumor than in spleen

			Median ratio of experiment versus input				
Intergenic region	**Genome position of peak signal**	**Arbitrary clone number**	**Spleen**	**Tumor (+)**	**Tumor (+)(-)(+)**	**Tumor (+)(-)(+)**	
			Lib-1	Lib-2	Lib-3	Lib-4	
STM0468 - STM0469	526177	**85**	0.9	**2.3**	**5.5**	**9.5**	TCAACTTGACGGTGCGCCAGCCACAGACTCAATCCTATCGGGAAA AGGACAGACAGGATAAGCACTCCCGTTACCAGGCTGACCAGATGT CGTGTTGTCACAGTGATGTCCTTATAAACACAGCGTAGAGAAAGTA TATCCGATCGTAAATCGCGCCCTCGAATGATAAAGCTATTTTATCG ATTTTACAGATTCAGGCGCCAGGCTAACGCGTTACGCCACGTTGCT TTTGCCGCCAGGAAGAGATCGTGAATGTTTACCGGTTGAAAAAGG AGCGTTGATAGCGTATTTTATTGTTATG
STM0474 - STM0475	529126	**86**	1.9	1.7	**3.2**	**2.6**	TATTGTTTGTGTAATCATTGGGTTAACGTTTTTTAGCTTTTCAGGCTA AAACAATAGACTCTGACAGGAGAAAATAGCCAGGAATATTCTTAAT ATTTCTTAATTAATGGCTGAATTAAGAAATGGCCAACTTTCCTAAGA AAAGCCTTTAACGCAGTAAGGATTATACCTTTTATTAATATGGCAAA AAATAATCAATCTAACAATAAGCGTATTTTATGATTTTTGCGTAAAA AAGGCCGCTTGCGCGGCCTTATCAACAGTGAGCAAATCAGCGATG TTCTGTCGAATGACTATGCTC
STM0580 - STM0581	638735	**87**	0.9	**3.2**	0.3	**8.5**	AAATAGCGAAACAATGTTCCTTCTGCAACACCTGCGTTACGCGCAA TCACCGCCGTTGAGGCGGCGATACCGGATTGCGCTATCGCCTGGG TTGCCGCTTCCAGTAATGCTTGTTTTTTGTCTTCACTCTTCGGACGA GCCACTACACGTTACCCTTATGTCTGGAAAAACATGATTGAATCAT GCCCGTTGTCGCGTCGCAACGGTGAATGTCAACCTTTGAAAAGTAC CTTGACGGCGTATCTTTGCTTTCTATAATGAGTGCTTACTCACTCAT AATCAAGGGCTGCCGCATGAAGTG
STM0844 - STM0845	914762	**10**	0.8	1.9	**5.8**	0.4	AGCCTTTGAGAAATACTACGGTACGGATACCGGGGCCATCGTGGG TAGAATAGCGCTGAATATTGAAGATCATAAACGGCCTCTCTTATTT CATATAAAGATTAAATTACTTTCGAATGAAAGCTATCTTGATGTGCG TCAACGAATGGAGAGGTTCTGACAAAGAGGCGTTAAATGAGGTAC AACATCACGGTTTGAGGTTGTGGTATGGCGTTTAAGATGATGCCGC GCTGCTTGAGCCGATCGTCAGTCGGAGCTTGGGTAAGCTGGCTTT GCGTCTGATGACAGTAATTATCTGTTG

STM0937 - STM0938	1014704	**11**	0.7	**4.2**	**6.5**	**10.3**	GCGTAGGAGCAGCCGTTTCCGGCTGGTGTACGGATGGTTTGTTCA CATTGCACACAAAACATGGTCACACCTTTTAAAGTTATATTTAATAT ACATGTTTAAGGTTATGCCTGTGAACAAAGGGATAAAAGGGATTTC TGCCATAATGTGCAGGGAGATTGATTTAGCGCAATTTTGGCGGCAG ATGCCTACCGCCAAAGAGGTATCAGGCCGAGAAGAACGCCATTAA GAGGGGGACCAGCAGGCTGAGGATAAAGCCATGTACGATAGCCG CCGGAACAATCTCTACGCCGCCGGAGCG
STM1382 - STM1383	1466034	**16**	0.7	**4.6**	**7.4**	**13.9**	TGAAGCATACCTGATTTCTGGAAATAGCGTAGATCGGAACGAATAG TCTCCTGGCTAACCTTATAAAGGTCTGAAAGTTTACTGACGCTAAC ACTATTATCCTTTATCAGTAAATTAATGATGGCATGACGTCTTTCTT CTTTAAACATATTGCCTCCGGGTAGTGAGTTGAATTGTATTTATGGC AATGTTGTCATGCGGTGAATTCAATCACAGATTATGCGGTCAACCG GAAGTAACCCCAAATGAATGTCAATAATCAGAAGCGCAGCCAATG TGTTAAATATTAATTGCTTACAGA
STM1529 - STM1530	1606103	**20**	1.9	**5.5**	**2.8**	**13**	TACACAAATGACCGTTTGCGCTATGTGATAATTAACCATAGTAAAA ATACACGAAGCGAAGAAGTGCTATTTCAGTAGTACTGATATTTTCA TAACGCTAATTTAAAAATAAATGTAAACGTAACAAATTATACACAA AAATAAGAAGGGCTGTGGCCTCAACTGACTGGATTATGATTCCGTC TTACCGAATGTCAGCCGAATGTTCAGTGCCATTCTCGCCCTGGCAT CCCCGACCGTAAGCCTGTTCTCTACTGGTAACCCCCTTGTTATTAC AGCAGAAAACAGGGCATATCATTGA
STM1807 - STM1808	1909051	**26**	1.2	1.6	**6.5**	**9.7**	TGCGCCGAACGCCAGTGGTCGTTTTTAACGCTGGAGATGCCGCAA TGGCTGTTGGGGATCTTTGCCGCTTACCTTGTGGTGGCGATAGCCG TCGTCATAGCCCAGGCATTTAAGCCTAAAAAACGCGACCTGTTCGG TCGTTGATACACACGCTCCTTCGGGAGCGTTTTTTTTGCCCGAAGC GTTGTTTGCCAGTGATTAAAAGGTGTATATTAAATACATCTTTTAAT CACCACATCAGGGAGATGTCTTATGTCCCACTTACGCATCCCGGCA AACTGGAAAGTTAAACGCTCTACCC
STM1914 - STM1915	2011503	**28**	0.9	**3.9**	**7.2**	**7.5**	GGATCTGCCCTTCTTCCCGCGCTTTTTCAAGTCGGTGGGGTGTGGG GGCTTCTGTTTTGTCGTCGTCGCTCTCTTCTGCCACGCAGCAAACC CTGGATAGATTGATAAGAGAGAATGATGCCAGAACCGCTTTACGC CAATAGGCAGAGTAAGCGGTAAAAAAGGCGGGGTTTATGGCGTTA ATAGAGATAGCCGGATACGATAAGAAAGTCTCGTATCCGGCCGGG TTGACGGATTCGAACCCGATAAGCGCAGCGCCATCAGGTCAAAAA AGCTTAAAAGCCAAGACTGTCCAGCAGGT
STM1996 - STM1997	2079476	**30**	1.2	**2.9**	**7.4**	**4**	GAATGGCTGAAAAATGCACAAACACATCTTTGCTGCCATCTTTAGG CGTAATGAAACCAAAGCCCTTTTCAGGGTTAAACCATTTTACTAAA CCAGTGATTTTCGTCGTCATAATATTGTTACCTTTCGAATGAGCCCT TGGGCAAAATGGCCTGAAGAAAATTATCAGAGAGAAAAAAACCTA AAGGAGATCTCAAGAGGAACAAATGATGAGAAATATTACAATCAC TACTTCAGATAAGTTTGTATCAAACCGCACAACCATTAACGCATGG TTAACTGAACATAGCAAGCTTTAGTT

STM2035 - STM2036	2114187	**31**	1.3	**5.9**	**4.7**	**8**	ACCACAAATGTGGCAAACCTGTTGGTTTACGTTATGGCTGTACGGC ACACCCATAACGACAATTAATAATGTGCTACGTTTTACATTTCTGTG AGCAATAGCCTGAGCGGTTGCTCATCTGACGTTAATCTACTCATCC TTACCGGTATATTGACGATAAAACGTATCGACAAAGCGTAATAAAA CTTATCTTTCCTGACACTGTACTTCATCACAAAAATAAAAACTGGTG CAGTTTATGCCCTAAATTTTATTATTTTGTTGCGCTATGACAATTTAT TGTTACACCAGATAAATTTTC
STM2261 - STM2262	2359663	**34**	0.6	**2.1**	**3.5**	**4.8**	CCTGGATGCAGGCGTCGCAACGCAGACAATGTGCGAGAAAATAGG TCGTTTCTCTGGCCCACGGCGGAAGAATCCCATTGCTGGCGTTGCG CCAACTGCCGGTCAACATGCTTCGACGGGATAAATCAACCATGAT ATCGCCCTTCCATAACGACACGCTTCCATAGGGAGTGAATACCAAT AAAAACCGTACAATTTATGAGTAGTTGTTTTTGTAAATAAGATATTT CAGGATGTGTAAGAGATGCATACCCCGATAGAGGTAAATGCTGTT GCCGGATCAAAAGAGTGCCGGGTAAAG
STM2309 - STM2310	2417301	**36**	0.6	**2.7**	**6.5**	**6.3**	TGAATAAAAGCAGGATTCTCTGCCGCCGCCAACGTGAGCGGCGTG GAACGGGAACCAGGGGCGATACAAACATGCCTGACGCCATGACG GGTTAAGGCTTCCAGGATGACCGCCGCCCAGCGCCGGTTAAATGC ACTTACTGACATGAGTTTGTCCGGTATCAATCATTGGGACTAAGTA TAAAGAGCTGCAAAAATGGATTATTGATATGGGTCGGGAATATGTG ACTCATTACGCATCCATCTGCAATAAGGTACGTAACCCGGCCGCTT TATTATCTATTTCCTGCCATTCCTGTTCC
STM3070 - STM3071	3233025	**44**	0.8	1.4	**2.8**	**3.1**	CGTTACGCCCGATGCGACCAAAGCCATTAATCGCTATGCGTACGG TCATAGGTCTCCTGCAAGGCTATCCCGATTCAGATGAGGCTGACAG AGTAATGCAGCTCATCGTCGAGTAAAACCTCACCTGTCGCAAACTG CGACTGATTGGTTAATTGTCGAACATTTAATTAACTGAAACGCTTCA GCTAGAATAAGCGAAACGGGGAATAAAAGGAATGTTTGTCCAGTC GAAGAAGACAGTTATCTGACCTGCATCACATTTCATGGCCGCTTAC GCTGCAATTTATTCCATATTTAAGAA
STM3106 - STM3107	3266543	**45**	1.1	**3.5**	**4.6**	**4.6**	TGATTTTGTTGCTGAATCACCACCGCCAGCGATCGTTCCGCCGGTC GCTAAGATGGTGATATTCGGTAAAGCGAACGCTGCGCCGCTGAAA CCCATTACCAGAGCAGCTAATGCCGTTTTCCTGAAAAACTCCATGT TATATCTCCAGTTATGTCAACTGGTCGCATTATCTCTATATTGCAGA CGAATAATGTGACGCCATACGATTAACCAGCGATATATATCCGACA GAGAGTATTTTTTAGAGATGGATAACAAAATGCAGGAAAAAACAG AATAAAAAGGCGCAGATACGATCTGC
STM3525 - STM3526	3688646	**55**	0.8	**3.8**	1.8	**5.6**	ACGCCTCTTCTACAGTGATACATTCAAATTGTTCCATGAATCGCTCT TTCATTATTGCCGGTGAAGCCAATTAAGGCATTTTATCGCCCAGTG TACGTTGACGGAGTAGCTTAGCGCCATAATGTTATACATATCACTC TAAAATGTTTTTTCGATGTTACCAATAGCGCGTTTCTTTGCTATTATG TTCGATAACGAACATTTTTGAACTTTAACGAAAGTGCAAGAGGGCA GCATGGAAACCAAAGATCTGATCGTGATAGGCGGGGGCATTAACG GTGCAGGCATCGCGGCTGATGCC

STM3880 - STM3881	4091492	**61**	0.9	**5.4**	0.1	**13.8**	GTATTTGCGTCTGCGTGGCAAGCTGTATTTGTTGTTGCAACGCAAC GCCCTGCGCGCGCCGGATCAGTTCGAGATCCCGCCTAACCGCGTG ATTGAGTTAGGTACGCAGGTCGAGATTTAACCTCCCATCAACATGC CGGGGGCCGCGTTGGCTTACCCGGCCTGGCCAATCCGTAGATTCC CACAAGATAATCGCCTGATTTCCGCTAGCGAAACGTTTCGACGGC GATCACAATTCTGTTACGTCATGATGGTTTTATGAACACATCCGGG GTTACACTGCGGCCAGCGAAACGTTTCG
STM4289 - STM4290	4530650	**71**	0.9	**2**	**8.3**	**10**	CATGTTGGTATCCTCAAAAAGTCAGCGGGGGCAAACGCGCCCAAA AATGGCAGATCGCCGAAAAAGGCCGCAATTATACACAAAATCCTT AGCGTTGTCGGGACTATTGCCGCTTTTATAAAAGGGTCTGCGCCAC GCCAGTCAGCAATGGTTTACACTCGAATAACCGCTTTTTTACTGTC ACCACAGCGCATTAGGGCGTCCTTATTTACACCTTTTGACCGAATT GACATATATGTGTGAAGTTGATCACATATTTAAACCCTGTTAGGGT AAAAAGGTCATTAACTGCCCATTCAGG
STM4418 - STM4419	4661108	**77**	0.8	**3.4**	**8.3**	**6**	CGATCTTATAGCTATTGAGAACTCTCGTTTCACAACCTATGTTTTAA TTTCAAAACGATCAATAATGAAACTTATGTTTTGTTATGGGTATCAC ATTTCGAATTTCATAATCCTGGCGTTTTTTATCGTTAAGATGCTGCG TTTTACGCAGTGCTCTCCTCTATCTTGATGAAGTTACTTGATTTTATT GATTTCGCGACAGTACCTGAACTCAATTTGTCAGGGGCCGTACTTT TTGTTCTTTCCTGGAACATCTCCATTTCGTGATCTTTTGCATGGAATT TTTCTTCTAATGAATGCA
STM4430 - STM4431	4674477	**78**	1.3	**6.1**	**5.6**	**8**	ACTACTGACTGCTTTATTCATTGACATATCCCCTAACAGAAGACGG TGTTATTTTTGCTCATACTAAGGTTTGGTGATTTCATTTTCAATAAAA ATGGAAATAATGTTTTCATTTATTGTTTGAACAAGATCACAGAAATG GCATTTCCGGGCAACGGGCATGATCGTTTTTTGTTGTGTTTTTTGTT TTAATTGATTGATTATAAATGTGTTATTTATTTTAAAATCGCATGGAA GATAAATTTCATTTTCATGAAAAATACGCCTGAATGTCGAAATTTTT TAACCGTTTTTTGATCTC

Table 2B Intergenic regions that induce higher GFP expression in tumor than in spleen (cont'd)

Arbitrary clone number	Cloned promoter orientation	5' gene	5' gene orientation	3' gene	3' gene orientation	Anaerobically induced	Stable / unstable GFP
85	+	*ylaB*	-	***rpmE2***	+		Unstable
86	-	***ybaJ***	-	*acrB*	-		Stable
87	-	**STM0580**	-	STM0581	+		Stable
10	-	***pflE***	-	*moeB*	-	Yes	Unstable
11	-	***hcp***	-	*ybjE*	-	Yes	Unstable
16	-	**orf408**	-	*ttrA*	-		stable
20	-	STM1529	+	STM1530	+		Stable
26	+	*dsbB*	+	**STM1808**	+		Stable
28	-	***flhB***	-	*cheZ*	-		Unstable
30	-	***cspB***	-	*umuC*	-		Stable
31	-	***cbiA***	-	*pocR*	-		Stable
34	-	***napF***	-	*eco*	+	Yes	Stable
36	-	***menD***	-	*menF*	-		Stable
44	-	***epd***	-	STM3071	+		Unstable
45	-	***ansB***	-	*yggN*	-	Yes	Stable
55	+	*glpE*	+	***glpD***	+		Stable
61	+	*kup*	+	***rbsD***	+		Stable
71	-	***phnA***	-	*proP*	+		Unstable
77	+	STM4418	-	**STM4419**	+		Stable
78	+	STM4430	-	**STM4431**	+		Stable

Table 3A. Regions that induce GFP expression in both tumor and spleen

Clone No.	Spleen	Tumor (+)	Tumor (+)(-)(+)	Tumor (+)(-)(+)	Genome position of peak signal	Genes and intergenic regions	5' gene	Function	5' gene orient.	cloned promoter orientation	
	lib1	lib2	lib3	lib4							
	Median of experiment versus input library										
Sequenced clones:											Sequnce
	9.42	2.94	1.48	15.51	711661	STM0648					
89	8.22	2.05	1.04	13.69	711724	IR STM0648 - STM0649	**leuS**	leucine tRNA synthetase	-	-	GAAGGATAGGGAAGCATCGACAGGCA GTAATACTTCTCTTTGCTCTCGTCTTCG GTCACTTCAAATGTGCGCTTCTCATCC CAGTGAAGCTGTACTTTGGATTCTATCT CTTCCGGGCGGTATTGCTCTTGCATGG CAGCCAGTAGTCCTGTTTTCGATACAG CTACAAATGTAGCTTTAGAGGTGGTGT TTAGATCCGCATAGCATAGCCCAAACA CGCACGTCAAAACAGGGGGTAGAACAT TTGTCGCGCCAGGCGTCCGTGAGGAG GTGACGCAAAATGCGACACGACTGAG GCAAA
	12.24	3.63	1.58	7.43	854765	STM0789					
8	12.94	4.32	1.62	7.43	854776	IR STM0789 - STM0790	hutC	histidine utilization repressor	+	+	CAAGAGTGCGCGTGGTTAACTATCAAA GAGCATGAGCCTTGTCTGCTCATTCGT CGTACAACCTGGTCCGCGTCGCGGATT GTTTCTCACGCCCGCTTACTTTTCCCC GGGTCGCGCTACCGGCTACAGGGACG ATTTATCTCCTGAGCGGACTGCTGCCG GAAAACGTGATTGCTGACACAATATAA CAAAATTGTATCATTTTTGTTAATTCTAT TCTTGTGCTTACTTGTATAGACAAGTAT ATGTCTGATTCTTATCTGTGGGTCTGC GGCGGTGCCTGATAGTGGCGTTTTAGC GT
	5.97	2.21	2.01	6.16	854930	STM0790					

12	3.55	2.26	1.48	6.75	1E+06	IR STM1055 - STM1056	**STM 1055**	Gifsy-2 prophage; homologue of msgA	-	-	GCTGTATTACTTCTGTAAACGCTGCCTA AACTATTTTGAATGTGTCTTAACATAAT ATACTCGCCGAATAGTAATTTTGTTAAT GTAATTATATACTACAGTGTGGATATTA ATACAATTCTTTTGTTGTTAATTATTATT TATGAAATTAATTAAAAGTGAATAAGTT AGAGGTGTTTGTTGGCCTTAAAATTACA TTTGTTGAGGGGGCTTATATGATATGTT TTTATTGTATTGTCGCATTTTTCTTAAGC TGAATCCGGATTTTGGGGAGGTGGCTA AATGTAAATGACGTGGTTTA
	3.37	4.00	1.33	12.90	1E+06	STM1056					
	14.51	3.69	4.70	15.31	1E+06	STM1264					
14	14.95	4.14	4.70	15.31	1E+06	IR STM1264 - STM1265	aadA	Aminoglycoside adenyltransferase	+	+	CAGTTGCCAGAAGATTATGCTGCCACG TTGCGTGCGGCGCAGCGTGAATATTTA GGTCTGGAGCAACAGGACTGGCATATT TTGCTGCCTGCGGTCGTACGCTTTGTG GATTTTGCCAAAGCGCACATCCCCACG CAGTTCACATAAGATGCCCCAGGACGT CTGTCAGGTTGCGCAAACGGCGTTCCT CAACTACTACTTAATAGGTTCTCATCGC TGAAGTAAGCAGATGATCTTATGCGGG CCATCGAATGGATATTCCCACATGGCT CTCGTTTTGTTGAGGTGGATATGACTG GTT
	14.98	5.19	4.38	12.05	1E+06	STM1265					
	6.70	7.16	4.44	21.25	2E+06	STM1481					
19	8.71	5.95	5.19	17.03	2E+06	IR STM1481 - STM1482	STM 1481	putative membrane transport protein	-	+	TAATGACGATTTTTAGACCATTGAGCGT GATGATCGGTTTTGCCATATCAGTCCC TGTTTTCTGATGCCGACACGAATAATAA TGTGATGTCGGTCGACCTGTTCTGGTT AAAATCAAACACTTCAGGTAAAGAAGT GAAAATATTTTGAGTTAATTCCTGGCTT ATGATACAAATCAGGCGTGTTCAACTA CCGAGGACAATTATCATCCGCGATGAC GAGAAGCAACACTGCGGATAATTGTAA TATTATGGACAATATGTTCAGCGCTTTT TTCTCCACGCAAACGCATCTTCACTCT
	6.11	3.79	0.21	11.96	2E+06	STM1686					

23	5.95	3.26	0.41	14.78	2E+06	IR STM1686 - STM1687	**pspE**	phage shock protein	-	-	ATTAATCGCGCCCTGAATATGCTCTCG CTGATATTGTTCCGGAATGCGGACATC TATCCAGTATTCTGCGGCATAAAGCGG CATGGCTATGAATAACGCTAACGCAAA TATTCCTTTTTTCAACATACTTCCGTCC TGACACGTAATGTATTTCGCACACACTA TACGCCAGAGCTTAACGAAATATTATGA CCAGACTCGCTATTTGTAACGCTGCGA AATTTTATTCGCCGCCTTACGAAGTACT GGCTCCAGCGCAAACGCCAGCAACATT TTTAGCGGACGACGGGCGACGGATTTT
	5.70	3.10	0.47	12.75	2E+06	STM1687					
	4.88	2.19	4.27	4.16	2E+06	STM1697					
24	11.13	4.14	5.28	9.30	2E+06	IR STM1697 - STM1698	**STM 1697**	putative Diguanylate cyclase/phosphodiesterase domain 2	-	-	ATCTTAACTCCCTGATAATGCGCTTTTA ACGCAAATCAATCAATAAAAACGATCAA TATATAAAAAATGATCGAAAAAACAATA TATGTTAACTTCATGATAACTTGCTAAT TTTATGTTTTGAGAATGTTCTTCTATTG CTATAAGGAAATTTACATACTACGCCGA ACAACGCTAATACGACGGCATGAGACC ATCCGTAAAGCCAGGTTTTTCTTGTCAG GCAGAGGGGAAAAATCAAGGCGAGTTA ATGTTGTTACACCATTGCGAGGCATTTC ACCCACTATGGCAGCGCGGCATC
25	11.89	5.62	3.76	13.35	2E+06	IR STM1805 - STM1806	**fadR**	negative regulator for fad regulon and positive activator of fabA (GntR family)	-	-	ATGACCATAGTGAGATTTCCATTACACA GCAAAACATAGTTGCACTCATCATACCA GACGGGCGTAACACCTGATAGCGGAC GCAATGAAGAAAAAGGGGATCAAGGCA CCATTTCTGATATCGCCTGCCAATATCG TTAAGGACTTGCTTGCATTCGTCGCGC TCGCTACTCTCTGTGTTTAAACATAAAA ACGCTATTTCATTTTTCTAGGTAAGGAA AAATTTCATGGAGATCTCATGGGGTCG CGCCATGTGGCGCAACTTTTTAGGCCA GTCGCCCGACTGGTACAAACTGGCACT
	12.08	3.58	3.13	11.54	2E+06	STM1806					

27	5.39	3.93	3.96	9.39	2E+06	IR STM1838 - STM1839	yobF	putative cytoplasmic protein	-	+	CTGAAAAGCCATTTTTCTACCATAGCTC AATAACTTCGCTTCTTCCAGTGCATCAA ATCACATTTAAAAGCTGTATTTTTCATAT CACTTTTTATGCTGAGTTATGCATAAAT TGTCACAATGATAAAAAACACCTTTTAA TCAAAATAATAGAAAAGAAAAGCGATTT TCGGCACCGCTTTTTGTGATGTTCTGC GTCTTTACAGAATGCCTTAAAATAATGA ACAAACAATGACAATCCATAAAGAGAG AGAAACGTTTCGCTTTTAATAGAGAATG AGCGGTATCACAAAAATGCCAT
32	10.42	8.43	4.63	14.61	2E+06	IR STM2122 - STM2123	**udk**	uridine/cytidine kinase	-	-	AAGGGGGGCGCCGAAACGCCAAACGC GGCAATTATAGGGATTTCAGCAGCGCG ATACCAGTCCGGCGCTATGCCACGGTG AATTTGTTGGCGGCGCATTCGACGTCG CGACGTAAAAGCGTTCAGTTTTAACGC GGGCAGCGGTTTTATCGACCCGTCTGG AGGAGGAATACGCCGGGAGCCACAAT TTATATTCAGCCAGCGTATAAATCATTA CGCGTTTATACTAGCATAATCACAGAGT AAACTGACGCGTCCGGTATTCCGCGAC GTTACCGGCGATTCGGATAGAGTGGTA ATGA
	8.12	6.36	3.56	11.86	2E+06	STM2123					
	14.55	10.26	7.87	17.67	2E+06	STM2182					
33	14.35	7.36	8.45	14.71	2E+06	IR STM2182 - STM2183	yohK	putative transmembrane protein	+	+	GCGCTGTGCCGAGCTGGATTACCAGG AAGGCGCGTTTAGCTCCCTGGCGCTG GTGATCTGCGGCATTATTACCTCGCTG GTAGCGCCCTTTTTGTTTCCGCTCATTC TGGCGGTAATGCGCTAACGACGGGAC AAAAGACCGGGTTAAAATTTGCGATAC GTCGCGCATTTTTCATTGAAGTTTCACA AGTTGCATAAGCAATGAGATTTAGATCA CATATTAAGACATAGCAGGCCCGTAAA CTACGGTTCCATTACATTGTTATGAGGC AACGCCATGCATCCACGTTTTCAAACT GCT
	11.03	8.54	7.69	12.87	2E+06	STM2183					

38	14.28	2.96	0.91	8.76	3E+06	IR STM2524 - STM2525	**yfgA**	paral putative membrane protein	-	-	ATTGCGCAGACGAACGCCGGTGGTTTG TGCTTCATTTTGGTCGTGCGTGGCTTC AGTATTCATTCGCTACAGCTACAGGTA CGTGTAAATTAGGATTCAGGCGCCGAC GAGCCGTAATGCCCGCCCACACCGCG AAACATCAGGTTAGTTAACCTTAGTCAG ACAGTATAAGCCTGTCAGGCCGCAGAT GACAAAACCGCTAAGACACAAGGCTAA ACTCTTGTTGCACCATTACATACTGCCT TAAAGTCGACAAAAACGCACCGTTATTA TTGACCAGACAAGTACAACGCCAGACA TT
	11.83	3.33	0.85	8.23	3E+06	STM2525					
	13.03	2.23	6.00	10.22	3E+06	STM2817					
40	6.85	4.27	7.12	9.22	3E+06	IR STM2817 - STM2818	luxS	quorum sensing protein, produces autoinducer - acyl-homoserine lactone-signaling molecules	-	+	TCCGGCATCACTTCTTTGTTCGGAATG CAAAAACGCAGATCAAACACGGTGATT GCGTCGCCATGCGGGGTGTTCATCGTT TTTGCAACCCGGACCGCCGGCGCTTG CATCCGGGTATGATCGACTGCGAAGCT ATCTAATAATGGCATTTAGTCACCTCCG ATAATTTTTTAAAAATAAACTGAACTCTT TGTTCCGGGGCGAGTCTGAGTATATGA AAGACGCGCATTTGTTATCATCATCCCT GTTTTCAGCGATGAAATTTTGGCCACTC CGTGAGTGGCCTTTTTCTTTTGGGTCA
	9.62	3.07	4.43	3.70	3E+06	STM3279					
49	9.70	3.07	4.43	4.57	3E+06	IR STM3279 - STM3280.S	**mtr**	HAAAP family, tryptophan-specific transport protein	-	-	AAAGACCAGCGCCGCCATCGACCAGA AGAACCACGCCCCGGACATGACCACC GGCAGGGAGAACATCCCCGCGCCAAT GATGGTGCCGCCGATAATCACCACGCC GCCAAGCAGCGAAGGTGACGTTTGGG TGGTGGTAAGTGTTGCCATTCAGCTCT CTCTCCAGTCATTTATAGTGTGACTATC TCTCAATACGCTGCACTGTACCAGTAC ACGAGTACAAAAGAAATAAAAAAAGCC CCGATTGTGACGATCGGGGCTGTATAT TTTACTTTACGCTGTGAATGCGCAGGT CAGCGTG
	8.14	2.72	5.09	7.11	4E+06	STM3441					

51	9.79	4.25	6.03	9.40	4E+06	IR STM3441 - STM3442	**rpsJ**	30S ribosomal subunit protein S10	-	-	TTCCGCGGTTGATTGATCGATCAGACG ATGATCAAACGCTTTCAGGCGGATACG GATTCTTTGGTTCTGCATGAGACCAGA GCTCCAATTATTTTATAAACGAAAATGA TTACTCCTCACACCCATTACGATTGATG GGAGAGTGTAACCGTTCTTACGTAGCT CCCCGATTGGGAGCATTGTTAAATAGC CAAATCGGCTATTCGAGGTTCAAATCG AACCTGCCGTCAATTACGACAAGCCCG CGCATTATACGTAAATCTCAGCCTGAC GCAAGTGTCGGATAGAAATTAAGCGCT TT
	8.53	3.07	1.15	9.96	4E+06	STM3499					
98	12.65	3.17	3.46	9.93	4E+06	IR STM3499 - STM3500	yhgE	putative inner membrane protein	-	+	AGCACAAGACGCCCTGCAGCAAACCG GTGAGCAACATCCCCCAGCGAGTAGTA TGTGAAAGCGCTACACTTTCCATGTCG TTATCCAGAATGATGAGAAAGCCGCAT TATTGCACCATCTGTTCACCGCCAGGC GTCGTCATGCATAATTCAGAAAAAAAC GCAGAGAGGTGAATCGATATTGTTAAT GTTGGTGTTACGTAACTTTCTTACATGA ATGCGATTACAGTCACATTATGTCGGT CAAAAACACTTCCTTTTAACGTTTTCAG AACATTTTCCACAACAAAAGTAGGTTTC CT
	2.45	3.73	12.35	19.22	4E+06	STM3500					
	6.69	2.72	5.18	8.20	4E+06	STM3568					
57	9.77	2.89	3.26	7.29	4E+06	IR STM3568 - STM3569	**rpoH**	sigma H (sigma 32) factor of RNA polymerase ; transcriptio n of heat shock proteins induced by cytoplasmi c stress	-	-	CCGTCAGCGAGCAACAACCGTGCCAAA GCCGATGAGCAACGAGAATATCACCCA CTCTTTTATCAGACAGTGATTTTATCCA CAAGTTCAATGTAACACTGTGCATAATT TGCACAAATCTTGTGACATAAAGATGAC GCGCGGGGAAGAGACAACAGGGACTC TTTCCCTGCGAACGGAAGCCCATTGCA GGGAAAGATTATACCACGATTTTATCAA TCGGGAGTAAAGTGACGTAAATGTTGC ACCGTGGCCAGCCAGGCGGCGATCCA GCCAATCATGGAACAGACCAGCAGCAG CA
	8.29	1.81	2.41	6.08	4E+06	STM3569					

58	11.88	3.48	0.80	7.56	4E+06	IR STM3621 - STM3622	**yhjR**	putative cytoplasmic protein	-	-	TATTTCTCACTGGCAGCATTACGCCCC GTCGTCAATACGGGAGAACGCGCATTT TTCATCTTTCCGTGACATCATTTATAAT GTGTAAAAATGCAAAGCGCAGAGTTAC AGGGCATCCTGCCGGGCAAATTGATTC ACATGCTAAATCTGATGCGTTTTAATTT CAATGTTAGGTTTATTTCTGTGCTTTCG CTAGTAAACTGATAAACAGTTAAAATAG TGACATGAGGGACACTGTGGACCCCGT ATTTTCTCTCGGCATCTCATCATTATGG GATGAACTGCGCCATATGCCAACCGG
	16.45	3.98	8.19	0.85	4E+06	STM3622					
59	7.64	2.84	0.85	8.98	4E+06	IR STM3624 - STM3624A	yhjU	putative inner membrane protein	+	+	AAACCGCGCCGGTTTCAGAAAACGCTA ATGCGGTGGTGATTCAGTACCAGGGTA AGCCCTACGTTCGTCTGAATGGCGGCG ACTGGGTGCCTTACCCGCAGTAAACCG AAAAAGGCCGCAAGGTTTCCCCTGCGG CCTGGTTCGGGCGCATGTTGCCATTAC GGCGGACAGACGCTCAAAACGCGTTA CTTCCTGTCACGTAGCCAGTTGACGAT CACACTGGCGATAATGCCAGCAATGAT CGGCGCTGCCAGATCGTGCCAGAAGA CCACGCCCAACTGCGTAAGCGTCATAT AGCCGC
60	7.89	2.21	5.33	8.90	4E+06	IR STM3838 - STM3839	**dnaA**	DNA replication initiator protein	-	-	**ATGATTGTTGGCGCACGTCGATAAGA CCCTGCATGAAGGGTGACGCACGAAC CGCTGTCTGCGGTTTTCACGGATCTTT CAAACGATCGCGACTTCACGCAGTCT GAAAAATTTCGTGTTCATGCCTGACCA GGATCGTTTGAAACGATCAGGACCGC GGATCATAGCCTAAACTGAGCAAGAG ATCTTCTGTTTCTCACAGATTCTTCCCT ATTTATCCACAGGACTTTCCAGGAAAG GATAAGTGTAATCGATCCTGGGGAAC TCCTGTACGCTTTCGCGCGCATATTGA AAAAATTAA**
	9.27	4.10	3.20	7.80	4E+06	STM3938					

100	9.27	4.10	2.88	8.41	4E+06	IR STM3938 - STM3939	hemC	porphobilinogen deaminase (hydroxymethylbilane synthase)	-	+	GTGTGACCATCGGCACCAGTTCTACCG TCAGTCCCGGATGGGTTGCCATCAATG CGTCTTTGACATAATGTGCCTGCCAAA GCGCAAGGGGACTTTGGCGTGTGGCA ATTCTTAAAACATTGTCTAACATGCTTG TTACCGTCATTATCAATCATTGACCATC CTAACATCCTTATAGAGAGTATGTTAGT TTTCCGGTCACCGTGAGTGAGAGGATA AGGCGCAGTGTCGTCAATGACAGTGAA TAATGACGAGAAACCGCCAGCCCGTAT TTAAGAATTTACACGCAGCGAACGGTG CT
	9.67	4.61	4.08	6.29	4E+06	STM3939					
63	11.21	8.20	5.10	11.30	4E+06	IR STM3967 - STM3968	dlhH	putative dienelactone hydrolase family	-	+	TAACAAACCACATTGCCTTAAAGCGGC TATCTTTTGTGCAATGCCTGGCGATATT GATTATTTATTGTGATGAACATCACTTT TTAATGGTAAGCGAGTGCAATTGTTTTA CGTCATAGTGATGGCTGTCACGAAAAT ATCTTTATGCCTTAGGTAAAGTGTCTCT TTGCTTCTTCTGACAAACCCGATTCACA GAGGAGTTTTATATGTCCAAGTCTGAT GTTTTTCATCTCGGCCTCACCAAAAAC GATTTACAAGGGGCCCAGCTCGCCATC GTCCCTGGCGATCCTGAGCGTGTGGA
	12.98	8.20	5.93	12.83	4E+06	STM3968					
66	9.91	4.92	5.25	10.47	4E+06	IR STM4087 - STM4088	glpF	MIP channel, glycerol diffusion	-	+	TGAATTGAATCATTTCATTAACCAATAT GTTAACACTTTTAAGTTATTGAATGAAT GTTACCAGGAGATGGATGAAAATTGCT GCAAACCGCGATCTACGCGGTATGTCG CTGGACAGCGAGAGCGGGGCTTCATA CAATCGACACTATATATTGTGCGCGTTT ACGTGAAGCGTCGCCTTGCAATTCAGG AGAGGTAAGATCATGTCTTTAGAAGTG TTTGAGAAACTGGAAGCAAAAGTACAG CAGGCGATTGACACCATCACCCTGTTA CAGATGGAAATTGAAGAGCTGAAAGAA AA
	9.91	3.66	4.69	10.65	4E+06	STM4088					

69	8.48	1.96	2.59	6.91	4E+06	IR STM4164 - STM4165	**thiC**	5'-phosphoryl-5-aminoimidazole = 4-amino-5-hydroxymethyl-2-methylpyrimidine-P	-	-	CAGCCTTTTCCACTTCATCCTTCGCGCT GCCTCTTCGTTGGCTTCGTCCGCTCAC TCCAGTCACTTACTTATGTAAGCTCCTG GAGATTCACCGACTTGCCGCCTTGACG CATCACGAACGCTTTTGTGGAAAATTA GCACTCCGACAAGATAACCGCCCCTCC GAAGAGGGGGCTGAAGTAAACTACCC GTTACTCGCGCAGAACTCAAGCGGGAC GTTTGACTCTGGCGCCGTCGTGCATCG CGTCAAACACCAGCATAATCAGCTTGT CTTCCAGCACAAAGCGGGCTTCCAGCG CTT
	16.14	4.52	2.44	17.65	4E+06	STM4165					
	9.06	5.41	2.57	13.59	5E+06	STM4335					
73	4.55	3.75	1.43	7.08	5E+06	IR STM4335 - STM4336	ecnA	putative entericidin A precursor	+	+	TTCGCGCCTCAATGATGAAACGCTTTAT CGGTCTTGTCGCGCTGGTTCTTCTTAC CAGCACATTATTAACGGCATGTAATACC GCCCGCGGCTTCGGCGAAGATATTCA GCATCTCGGCCACGCCATCTCCCGTGC AGCCAGCTAATCGCTTCTCGTCTTCCT AAAATTAGTCGATCGCCCATCATTTTCT GGGATGTTGTCTATTATTAAGTTGCTAT ACACAAACAACATTGGCTAGAAAAGGA AGACATTATGGTTAAAAAGACAATTGCA GCGATCTTTTCTGTTTTGGTACTTTCC
	3.12	2.34	0.87	3.98	5E+06	STM4336					
	10.88	3.11	4.71	12.55	5E+06	STM4399					
75	17.04	4.02	5.83	15.54	5E+06	IR STM4399 - STM4400	**ytfE**	putative cell morphogenesis	-	-	TTTCCGCCGCAGCAGTAATCCATATCG TACTGGCGAAACAGCGCCGATGCGCG GGGAATAGAGAGCGCCAGTTCGCCTAA AGGTTGATCGCGATAAGCCATAGCCGT TACCTCATTTGCAATAATATAAGTTGTA TTTTAAATGCATCTTTAAGGCGAAGCTA TAACTCTTTCGGGGTGCGTATAATTTAA GCGAGTATGAAATTAGCGTTCCGTGAC CGGAACGACGGTCGCTTTTTCCGGTTT CGCTCTCACGGCAATGACCACGCCCG CCACCAGGAGCGCAATGCCGCTTAAC GTCA
	14.72	4.99	5.83	17.37	5E+06	STM4400					

76	12.10	8.37	0.91	15.76	5E+06	IR STM4405 - STM4406	ytfJ	putative transcriptional regulator	-	+	GTGATCCGACCACTTTGGGCCGATAGT TAATCATATGTGCGATTGATGCTTTTTC CCGCAAAGGGGATGCCAGTTTGCGGG CGGGCGCACACTTCCTGTGAAAAATGA AGGCATATACTGAGAAAAATGAGCTGA TGTTTAGATAATTCTGAATAACTGTAAT CAAAAGGTAAATATACTTATGCACACTG GAAACGACGTAGATATGGTCTATAGTC ATATGGCATTAAAATTTGCGCCTTAAAA CTGTTGGGCCGATTGTGGCATCGCAAG GGCGTAATACTCTGCAGGAGACAACAA T
	11.07	9.07	0.91	14.42	5E+06	STM4406					
	7.73	4.88	4.40	7.19	5E+06	STM4484					
82	7.87	4.97	4.70	7.43	5E+06	IR STM4484 - STM4485	**idnD**	L-idonate 5-dehydrogenase	-	-	GATAATAATGTAAGTCAGACCCACAAAT GCCGCCACGGGTAATTTGTACGAGAGT TCCTTTATTATTCCATTCAATATTTTGTT CCGTAACGGCAACAGCACGCTTACCCG CAACAACGCAGGATTGAGTTTTTACTTC CATAAATTCCTCACTGGTCAGGTAGTTA CCCTGAACGCATTTAAGCGGTTTTATTT GTCACTATTTGTGACTTATGTCACGCTG GAAAATTGTTACACTACAATGTTACGCA TAACGTGATGTGCCTTAGAGTTCTTCTC TATGGAAATTAAAAAACGTGAA
	4.40	3.55	6.66	4.67	5E+06	STM4485					
102	6.83	4.51	1.52	4.48	5E+06	IR STM4551 - STM4552	**STM 4551**	putative diguanylate cyclase/phosphodiesterase domain 1	-	-	ATACACGGAATCGGGCGCCAACATGAA AATAACGTATGAGAAAAGGTCGCCTAA AGCGAGGTGTTGTTGTTTTTACGTTAAC AGTCGGACAATTTATCACCTTACTGAAT ACGTGTCATCAACCGTTAAGTAAAACTC ATCTCTTTAGCTTTCTCCCTGGCTGACA AATGAGAAAATATATCATATGATATTGG TTATCATTATCAATTCCAGAGGTGAAAC CATGTTGCAGCGGACGTTAGGCAGCG GATGGGGCGTATTATTGCCTGGAGTGA TTATCGTTGGACTGGCGTTTATCGGC
	8.88	3.83	1.44	4.96	5E+06	STM4552					
	5.54	5.79	4.40	14.79	5E+06	STM4566					

83	10.24	5.19	8.33	14.49	5E+06	IR STM4566 - STM4567	yjjI	putative cytoplasmic protein	-	+	CGCTGCTGGAGCGCAGTTTCGCATGA GGCAGGCATCTTCGTTTCCTCTTTATG CCGGGACGATGCGCTATTGTAGAAAAT GGCGGCAAACCGACTTTGATCCTGATG CGCTTATCGCTCGAAGAACAGACGGTG ACGGCGGGATAATTTGATTCAGATCTC ATTACAGTAATGCAAATTTGTACGTAGT TTTCATTAACTGTGATGTATATCGAAGT GTAATCGCGAGTGAATGTTAGAATATTA ACAGACTCGCAAGGTGAAATTTTATAC GGCAATGCCGTTGGAGAATGTCATGAC TG
	8.07	5.72	5.32	11.30	5E+06	STM4567					
Supported by array data only:											
	7.53	3.93	3.12	16.10	39114	PSLT047					
	6.23	9.42	4.09	21.40	39436	IR PSLT047 - PSLT048	PSLT 047	putative cytoplasmic protein	-		TTCTACCGGATGGTTGAGCACGTTCAT TTCATAAAATGATGCAAATTCGCCCCTG TCAAACACGGCGCCGAAATCGGCTACC GCTTTCCACACTTCGCCGCGATCGACA TTGACAAAGCCTTTATTCCAGTCGCCAT ATCCGAAGCTAAGTTTACCGTATACGC GTTTCAATTCCGCTGCCTGGCCATTAA AGCAAGAGAAAAGAACACATGCGGCGA GTAGACTATTAATATATTTCTTATTTTTC ATGCTCAACTCCATGAGGTAAAAACAC AGTGAAATGTTGTGTAAAGAAGCGAAT
	4.20	5.90	3.12	12.13	108368	IR STM0093 - STM0094	imp	Organic solvent tolerance protein	-		GGTCACAGCCTAACTTACTCATCTTCG CTGCGCCAGTGTTAATCCTGCCGTTTA GCGTCTGTGGTGTTAGGCACGGCATTG AATGACAGGTATGATAATGCAAATTATA GGCGATGTCCCACAATTGACCGTAGCC TTCATTTGCAGAAAAGCACCTTATTTTG TGGGAGATAGCCTCACCGATAGCGTAA CGTTTTGGGGAGTCTATGCAGTACTGG GGAAAGATAATTGGCGTCGCCGTAGCC CTGATGATGGGCGGCGGCTTTTGGGG CGTGGTCCTGGGTCTGCTGGTGGGCC ATAT
	7.78	6.97	5.53	15.14	108588	STM0094					

	16.16	4.53	1.45	6.75	230588	IR STM0194 - STM0195	fhuB	ABC superfamily (membrane), hydroxamate-dependent iron uptake	+		TAAATAAAAAACGCTTGTCTTTGGGTTT TTAATGGAAAATACTTCACCGCGCCTAA GGGATGTTATTTATTAACGTGTTGTTTG CTTCTTTTGAATGTTGCATCGGCAATTT CATAACTCGTCATATAATATATCTAC TAATATAAACATGGGGTATTGAGTATAA CTCTGTGTGAATAGCGTAAAAATACTCA CCAACTTTTAATAAGGATGAAAAATGAA TACAGCAGTAAAAGCTGCGGTTGCTGC CGCACTGGTTATGGGTGTTTCCAGCTT TGCCAATGCTGCGGGCAGTAATA
	16.16	4.05	1.60	7.30	230618	STM0195					
	5.06	3.61	3.18	11.78	256949	STM0218					
	5.06	3.81	3.87	10.76	257001	IR STM0218 - STM0219	pyrH	uridylate kinase	+		GCTGGATAAAGAGCTGAAAGTGATGGA TCTGGCGGCGTTCACGCTGGCTCGTG ACCACAAACTGCCGATTCGTGTTTTCAA CATGAACAAACCGGGCGCGCTGCGTC GTGTGGTGATGGGCGAAAAAGAAGGG ACGTTAATCACGGAATAATTCCCGTGA GCGCCAAATACGGGTAAGATTCTGTTC TATTGACGGGTCTTATTACCTGGCAGA AATTAAACGAGACTATACTTAGCACATC TTTATATTGTGTGACCGTCTGGTCTGAC TGAGACTAGTTTTCAAGGATTCGTAAC GTGA
	13.58	3.14	2.83	10.90	258882	STM0220					
	9.50	3.85	3.09	6.86	259045	IR STM0220 - STM0221	dxr	1-deoxy-D-xylulose 5-phosphate reductoisomerase	+		GATTCGTTTTACCGATATCGCCGGGCT CAATTTAGCGGTGCTGGAGAGGATGGA TTTACAGGAACCGGCAAGCGTTGAGGA CGTATTGCAGGTTGACGCCATCGCGCG TGAAGTAGCCAGAAAACAAGTGATACG GCTCTCACGCTGACGATTATCCCGCGA CAGAAGATCGTGCTATTTGTTAGCGTT GGGCTTCGGTGATATAGTCTGCGCCAC CTGATCGCAGGTTTTTGGCTTTTTTCGG TCAGGTTAGCCGTGGTTTTACACGGCT TTTTTGTGGATACACAAAATCATTCAGG AC
	9.06	3.02	0.27	4.57	280369	STM0238					

	9.81	4.01	0.73	7.77	280632	IR STM0238 - STM0239	yaeP	putative cytoplasmic protein	-		AATATTTTTCCACATGCCCTCCTGTCAG CATTCTGACTTAACCGTGGATGCAAGT CTAAGCCTACGAAGTTAAATCTTGTTTA GCAAGGTGACTATACCATACTCATTTG CGCAATATCAGCGCCTGACGCGAGTG GGTAAAAGATTCGTTAACAGCCTTTTAG CGCGGTTTTCGCTACAATGGGCGCCTG ATTCGAAAGGAGTTTTCTCATGGCGCT TAAAGCGACAATTTATAAAGCCGTCGT CAATGTGGCTGACCTTGATCGCAACCG GTTTCTGGATGCGGCATTGACGCTGGC GC
	9.19	4.19	0.72	7.77	280644	STM0239					
	21.74	9.05	6.68	14.14	350300	STM0306					
	23.71	2.23	3.60	6.98	350713	IR STM0306 - STM0307	STM 0306	homologue of sapA	-		GACCAGGCTACCACAAGGGGAATGAT GCAGACTGCGAAAAAGTTTTTCATTTCA GAACCTGCCTTAATATTGGGCTAAAAG ACAAGTTTCACGGTATAGGGTGTGATA TAACGATTACATAAACGAAGCCCAAAAA ACGGTCTATTGTAACGCTGGGTTTTCT GTAAGCGGGTAAAAAATGAGATGAAGA TTTTAAATAACAATACGATAATCGTCGG TATGGAAATCCATCTCCTCGCCAAATTG CCCCACGTACGGTTTCACTTCTACGTT ATGTAACGGGTAGTGTGAGATGGAGCG A
	18.23	3.38	2.66	8.07	350910	STM0307					
	4.50	3.64	1.20	6.94	385496	IR STM0340 - STM0341	stbA	putative fimbriae; major subunit	-		AAACAGTATAATTAGTCTTACTTTTTTCT TACTTTTGGCCTTTCAGAAGTTTCCTGA GTTTGCGTTAAGGTAAAGAAAAGTGTT CAGATTTACCTATAACTGTTTGATTTGT AATGTGTAGGTAATACTTGTGTCAATTA TTGTTTACTATAAGTGAGACTTATAAGT TAAACTCAGGTTAATTAGGGGGCTGAA TTCTTTTTTGAGCATGATAATATGTCGT CTGAATGATGGATGCAGTTACCTTTAG GATTGTCATGAATGAAACTATATTTTTA CTTGATAAGCGTGTTGTATTTGA
	4.42	3.55	1.12	6.31	385529	STM0341					
	6.92	7.96	4.23	12.59	386588	STM0342					

	7.27	7.41	4.09	11.40	386656	IR STM0342 - STM0343	STM 0342	putative periplasmic protein	+		AATCCGGCAGGATTACCCTACACTACG ATGTTACTACCGATACGAAAGAGAAAC GGCTTTTTTTCGTGATATCTGCATCAGC AAACTGCGCAGAACGGGTATGAAAACA TTTACTTTTAAAGTCAATTCAGTTAAGA CTTTTGAGTCTGATACTGCTGGCGATTT GTTTTCCTGGTTGAGACTGTTACAGCC TGGTACGATTAATGAGTTAAAGATGGT CAAAATTGGGAAAAATACCTACATGTTT TCGCTTAATCGACATTTGTATAATGTGT GTACCACCAGTAGTAACGTTGAGTTG
	2.14	2.18	0.75	4.10	450515	STM0396					
	8.70	2.17	1.65	3.75	450651	IR STM0396 - STM0397	sbcD	ATP-dependent dsDNA exonuclease	-		AAAGCCTGATGCTCCGCGGCGCGGCT TTTACTGTAGAAATTTTGTCCCAGATGC CAGTCAGAGGTGTGGAGGATGCGCAT AATTGTTCCATGCAAAAAAAGCGTGAA CGGGATTATACACGTCATCCCTTCCATT TTTGGGCGCAATTTACCGCCGGTACAC GGTAATGCATGGTTTCACCGGTGTCAT AAATCATCAACATGCTGTCAATGCCGC CTTTTTTTTTCATAAATCTGTCATAAATC TGACGCATAATGGCGCGGCATTGATAA CTAACGACTAACAGGGCAAATTATGGC GA
	12.04	5.51	3.16	0.46	450902	STM0397					
	11.06	4.11	2.66	12.37	508340	STM0451					
	11.06	4.38	2.82	12.37	508386	IR STM0451 - STM0452	hupB	DNA-binding protein HU-beta, NS1 (HU-1)	+		GGTAGGCTTTGGTACTTTTGCTGTTAAA GAGCGTGCTGCCCGTACTGGTCGCAA CCCGCAAACAGGTAAAGAGATCACCAT CGCCGCTGCCAAAGTGCCGAGTTTCC GTGCAGGTAAAGCGCTGAAAGACGCG GTAAACTAAGCGTGATCCCCTCGGGGG ATGTGACAAAGTACAAGGGCGCATCAA CTGATGTGCCTTTTTTATTGGCGATTCG GGACTTTCTGTGCGTTGCGGGCTGACA ATTGCCCTCGTTTCTTGTCACAATAGGC TTTTGTGCGCCGCGTTCAGAAAATGCG ATGC
	7.10	8.00	0.37	10.82	522980	STM0464					

	5.77	4.81	0.36	9.15	523177	IR STM0464 - STM0465	tesB	acyl-CoA thioesterase II	-		CTGACCGCCAAATACCTGGCGCAGCC CTAAGTCTTCACTTTGGCCCCGAAAGA GTCCTTCTTCAATTTTTTCCAGATTCAA TAATGTCAGCAAATTATTCAGTGTCTGA CTCATACATACTCTCCAGGTGACAACG ATGCCGAAGCGAGGTAGGGCAGAGTA TAACGCAATTTTGCAAGTGGTCCGATG GGTACAAAAGTCTGAATAACAGACCAA TTCCAGGCAAAAATGAGTGACATGTGC CACACTTAATCACGTTATGTTTCTGTTA ACCACTCTTCCGGCGGGGGGAAAGGC CTGC
	5.75	6.67	6.06	9.71	533588	STM0476					
	6.79	6.13	6.93	8.40	533647	IR STM0476 - STM0477	acrA	acridine efflux pump	-		TCTGGCATCTGCTGGCCGCCTTGCTGG TCCTGTTTGTCGTCACATCCTGTTAGC GCTAAGCTGCCTGAGAGCATCAGAACG ACCGCCAGAGGCGTTAACCCTCTGTTT TTGTTCATATGTAAACCTCGAGTGTCCG ATTTCAAATTGGTCAATGGTCAAAGGTC CTTAAACCCATTGCTGCGTTTATATTAT CGTCGTGCTATGGTACATACATCCATA AATGTATGTAAATCTAACGCCTGTAAAT TCACCGACATATGGCACGAAAAACCAA ACAACAAGCGCTGGAGACACGACAACA
	7.34	5.05	4.44	12.10	534374	STM0477					
	7.30	6.03	4.23	13.57	534417	IR STM0477 - STM0478	acrR	acrAB operon repressor (TetR/AcrR family)	+		TCAGGGCTCATGGAAAACTGGTTATTT GCTCCGCAATCGTTTGATTTAAAAAAAG AAGCTCGCGCCTACGTCACGATCCTGC TGGAGATGTATCAATTGTGTCCGACGC TGCGCGCGTCGACGGTCAACGGCTCC CCCTGATAATATTCCAGGAAAACTCCT GGACATTTTCTGTGTCGCTATTCTGTTT GTTACAGGCGTGATATTCTTGCGACTC AATTATTTCCGGTCTGCTTGCCGGTTCA GACACTTCATTCTCATGACTATGTTGCA GCTTTATAAACGTTCACAGCATTTTGTT
	5.99	5.29	3.53	12.94	534476	STM0478					
	2.86	2.34	0.61	8.04	598959	STM0536					

	3.16	3.01	0.64	10.18	598994	IR STM0536 - STM0537	ppiB	peptidyl-prolyl cis-trans isomerase B (rotamase B)	-		ATGGTGTTGTTGTAAAAACCTTCGCGG CAGTAGTCCAGGAAGTTTTTAACTGTTT CAGGCGCTTTATCATCAAAGGTTTTGAT TACGATATCGCCGTGATTAGTGTGGAA AGTAACCATTTTTGCATCCTGTTCCAAG AGAGTGGTGCTTTAGCCCGCAATGGG GCACATATAGGGGCTTGTTATAGCATA ACCGTAAGCTGCGATCACCTTGCAAAG TGTGCTGCTTCGATTACGAATAATATGT ATCATACGGAGATTATTACCCACACAC GTCTATACGGAATCTTCGATGTTAAAAA
	2.62	2.98	0.54	7.94	599106	STM0537					
	6.23	2.91	0.44	8.74	649485	IR STM0588 - STM0589	entF	enterobactin synthetase, component F (nonribosomal peptide synthetase)	+		ATTAATAAATAACGGGCGTTGTTTCTGC CTTTAACAAATTAAATCCTGAAACCCAT AATAATTACTAATTATTATGGGTTTTTTA TTGCAACTATTAATTCTTTTAACATAAGT GATACATGCTACAGGCAAGTTTAATTCC GAATATTTAGCTTTTCGGGCACTGGCG CGTAAAGATTGTTTCGGATAATTCTGAC TTGCTGTTAGAATCTCTGACAGGAATGT GTTCTTTCATTGGATAAAGTTTTCAGGT CATACGGCATGCCATCTCTTAATGTAAA ACAAGAAAAAAATCAGTCAT
	5.62	2.58	0.36	7.48	649550	STM0589					
	8.75	5.12	3.69	15.76	704993	IR STM0642 - STM0643	ybeB	putative ACR, homolog of plant lojap protein	-		ACGCCGTGTAGTATACCTGAATCAGCG GCGATACCGGGACTTATGTCGCCGGAT CGGCGTTTAAAACCAGATTATCATCCC ATCCCACGTCACAGAAAGCATCGCCAT TTTTGTAAAACAATTTCTGCAAAGCTCT GCAAGGTGAAAAAAGCCTGGCTGCGG AGAATAACAGCCTGTCGGGGGCTGTCA ATGGGCGAAACCGCTGCGGCGAGAAA AAACGGAAAATTCATCACTCAGGCCGC CAGACGGCACGACTATTTAATACTTTCA GGGTGGCGAACCCTTCGCATATGTCGA TTGC
	9.05	6.18	3.69	17.29	705024	STM0643					

	11.63	6.24	8.80	8.43	766043	IR STM0701 - STM0702	speF	ornithine decarboxylase isozyme, inducible	-		CAATAGACCTGAATGACATAAGGGTCG GAAAGACCTGTATGCTGAAGTACCCGT AGCAGAAAAACTACCGGGCATTAAAGA AATGAAAGTCGAAACTATTGCGGTGGG CAAACATCATAATATGCGTTGTCCGCCT TATATGGGGCATAAAACGATTATTATTT TCCATTTTGAGGTCCTTTCATTGATTTA TTGAAAGCATGGATATTTTATCCAGGAA GCGCCAGCAATCTGTGAACCAGATCAA CAAAAAACGATCATTTGAAAAATAATTA GTCGGCGATTATGCATATCGTGCTGT
	17.22	6.49	7.28	11.13	826178	STM0762					
	12.09	3.34	5.14	8.39	826326	IR STM0762 - STM0763	STM 0762	fumarate hydratase, alpha subunit	-		TAATGGTTTCCTTGCCGATCTTTGACTC TTCTTTATCATATGCTTTACGAAAAGAA CACATGAGATTATCATCCAGTTCATAAC AAGCTTTTTTTACAAGTTTTTCGATAATC GGAATGATAATTTCTGTATTTAATATAC GACTCATACTCCCTCCAGTGCTATGTT GCATTGTTTTATCCATTGATCACATTTT CATGATATTCGTATTCATTGTAGGAAGG AAATATGTTATTTTTATTAAATGATAAAT TTTATTTATAGTAGTGGAAAATAGATGG AAATTAGACAATTAGAATAT
	2.29	5.25	4.55	10.15	901671	STM0834					
	7.34	4.71	0.34	5.13	902051	IR STM0834 - STM0835	ybiP	putative Integral membrane protein	-		AATGGGCGCCATTTCCGTTGAGGATGC AAAATAAAGCGGCGTACCGCACCGCC GCGTTATTTCGTGGAAGGGTTATCCTG CTCCGGTTTGCCGTTGATCATATCGCA CAACATAGAGAGCAGCATTAACCGGAC TTTAAAGGGAGAGTGACTGAACACGCG TATACACCTCTTAAATTCGTTCATATAA ACCTCCTGATGTTTCTATCCCATCGATC CGTGAGGGATGTCTGCATTACATACAG ATATAGCACAGGCTATGTTTTATAGCTA TTGCTAAAACGTTAATTTTTTGTGCCCA G
					902276	STM0835					

	14.20	5.38	2.63	8.80	932960	IR STM0859 - STM0860	STM 0859	putative transcriptional regulator, LysR family	-		CTACCAGATGCGGCAGACATGTAAGTT TTTTCCGCTCCACGTGTTATGCTCCCTT CTTCACTGATAGCAAGGAATAATTTTAA ATCTTTTATATCAAAGTGCATCGTTGTG GCTCATAATTAACGTATAATACAGTGTG CTGCTTTTTTATAGACTCAGTCAGACTG AGTATTTCGGCCTATCCGAATTCCTGTC ACGTCGAGATAACTACAAAATGTAGGC TGACGGTGTCACCGCCCTACCATGATC CGGGGCGGATCTGGTAGGACGCTGGT GACCGCTGACAGGGGGTCAGGTCAGA
	13.76	7.84	2.74	10.87	933137	STM0860					
	5.18	4.54	0.74	9.72	1E+06	STM0943					
	8.61	7.82	1.91	22.11	1E+06	IR STM0943 - STM0944	cspD	similar to CspA but not cold shock induced	-		TCAGGCGAGGCGTCAAGCATCAGGCA GGGGGGATCGGGTAAAAATGAATCAAA AATTTGAAGCAGTTAACGCTATTGCCG GGAATGTGACAGATGTCGCGGATGGTA CTGATAGATGTTAGTTATCTATCAATTG AGGTAGATTGATTGTGTGCATAGACTC TGGTCAGCGGCAGATTTTCCTGCCGAC AACTGTAACCGATAATGACGACTGACA ATGGGTAAGACGAACGATTGGCTGGAT TTTGACCAGTTGGTGGAAGATAGCGTG CGCGACGCGCTAAAACCGCCATCTATG TATA
	8.61	3.76	1.91	21.37	1E+06	STM0944					
	3.93	4.39	1.02	11.82	1E+06	STM0946					
	2.43	3.12	0.93	4.12	1E+06	IR STM0946 - STM0947	tnpA _1	IS200 transposase	+		TATCTGAAGGGTAAAAGTAGTCTGATG CTTTACGAGCAGTTTGGGGATCTAAAA TTCAAATACAGGAACAGGGAGTTCTGG TGCAGAGGGTACTATGTCGATACGGTG GGTAAGAACACGGCGAAGATACAGGA CTACATAAAGCACCAGCTTGAAGAGGA TAAAATGGGTGAGCAATTATCGATCCC GTATCCGGGCAGCCCGTTTACGGGCC GTAAGTAACGAAGTTTGATGCAAATGT CAGATCGTATGCGCCTGTTAGGGCGC GGCTGGTAAGAGAGCCTTATAGGCGCA TCTGAAA

	4.71	5.27	1.14	8.16	1E+06	IR STM0958 - STM0959	trxB	thioredoxin reductase	-		TGTAGGGAATTTACAGACGTAAAAAAA GAGCATAACGATTTTGTTAACAATATGT GTAATAGCATGAACCGATGAACGGCCG CGACAGCGACGTTATCATCACAAACTT TAATTAAAATCGGTAACTTATAAGGTGA CGAAATGACAGTTTACCGCCCTCTCTA ATGAATAACTGGCATGTTGTACTAAAAA TCGATGTTTTGCTTTGACAATCACCTGC TGTTTTGCGAAAACATTCGAGGAAGAA AAAACTGTGTTATGTATGTGCTGCATAA TCATGCATGTAAATACCATGTTTACC
	5.19	7.82	4.90	14.40	1E+06	STM0962					
	4.40	9.12	3.63	14.04	1E+06	IR STM0962 - STM0963	ycaJ	paral putative polynucleotide enzyme	+		GCCCCACAAAACGCTACCGCTAGTGTA AACGTTGCGGTAAGGTTATCTCTAAATA TGATGCTCCAGGTATCATGGCGTTGAT GATGAATCTCGTTATGCCTGATAGCAC GTTGCTTATGAGGTCCGCGGGTATAGC GCAATGGATGCGTTGTTGCTGTCGTCG GTCTGGTAAGGCGAAAACGTCGCTATT ACGTAAACGCGGTTTACGTTCATCAATA CAATCAGAGGCGATCATCAATTGATCG CGTTTCCTTTTATTATTCGATAAGCACA GGATAAGCATGCTCGATCCCAATCTGC T
	19.39	4.17	2.54	0.28	1E+06	STM0974					
	4.76	3.09	4.28	4.25	1E+06	IR STM0974 - STM0975	focA	putative FNT family, formate transporter (formate channel 1)	-		CCTGGCTTATAGGCCCGTAAGTCGCAT GGCTTTTATGCAATTACGGTGTAACTTT TTGATTATCCTAATAAAAATAAATTTTAA AAATTATAAATAGAGTTGAATTTTTTCCT GACTCCTCCTGCTGCACGGTTAATTAA TATGGAGTAATCAACAAATAAAGTAACA TCACTATGTCAATTAATTTAATATCAACA ACCAATATTTAACCTTGTTATTACATTTT TCGCCGTTTAGCGAAAATAAATAAAAC GGGGCCGCAAAGGCGCCCCGTAATAT AACGCAGCCGAGAGGGTAAACC
	6.85	5.88	0.71	8.94	1E+06	STM1000					

	9.45	5.61	0.38	11.22	1E+06	IR STM1000 - STM1001	asnS	asparagine tRNA synthetase	-		CACCCATCCGCGCACGGTGACTTCTTG GTCAACGGCTACGCGGCCCTGGAGTA CGTCGGCTACAGGCACAACGCTCATAA TATTCTCTCTAGTTAATAGTCGGAAAAA ATAAACACTTGTCCACCCGAAATGGGG GTATTCCTATGTTACCTGGCATCTGCAA TCAGACAAGCAGAAATCGCATCTGGAA GCAGGTTTTCAGAAAGAAACCTGTAAA AAGTTCGCACCTGCTCGCGAACCATTG AGAATTTAGGCTGGTTTTGCAAGCTTTG CGCACGTTACTCGATCAGGACGCGCAT CT
	6.14	5.36	0.30	7.51	1E+06	STM1001					
	3.99	4.52	0.27	9.86	1E+06	IR STM1019 - STM1020	STM 1019	Gifsy-2 prophage	+		TTTGATGCTGCTGCCGACAATTTTTAAC CGCGTCCGTGTGTCGCTCAGGGGGGT TACGTGGCAGAGGGAGTCCTATCAGAT CTTGCTGATAATTTGCGGGTGACTATAA CTGATGCTAAGGGAATAGAACTTTTGT CTTTTAGACTTGCATCAGGTGATCGCTA TATCCTATCAACCCAAAACGGTTCTGTA ACAAACCGAAAGCTATCAAGAGATGAT TTGTACTGGTCTAAGGATACCATTATGG AAGTTGTCAGAGAGATGGGCTCTAATA ATTGACTTAACAATAAGCACGCAATCA
	7.78	2.62	2.75	11.74	1E+06	STM1070					
	13.38	4.07	4.15	9.95	1E+06	IR STM1070 - STM1071	omp A	putative hydrogenase, membrane component	-		GTCTTTTTCATTTTTTGCGCCTCGTTAT CATCCAAAATACGCCATGAATATCTCCA ACGAGATAACACGGTTAAATCCTTCAC CGGGGGATCTGCTCAATAGTTACTCTA CCGATATCTACGGCTTATGCTGAGCAC CCCTGGCGATGTAAAGTCTACAACGTA GTTGGAAACTTACAAGTGTGAACTCCG TCAGACATGTGAAAAAAACATGACGGA TATACACATCATTTAACAGTTTCAGATG ATAAATCGTACAGCAAAAATTGCGGAA ACCGCTTCTGACAAGCGTTCTCGCAAA A
	8.17	1.31	2.77	2.51	1E+06	STM1094					

	8.43	2.49	3.03	11.31	1E+06	IR STM1094 - STM1095	pipD	Pathogenicity island encoded protein: SPI3	-		TAATGAAGGAGCCGTCAGCCGAAGCCTGATTGCCTACCAAAAGGGTAGTACAGGCGATGACTTTACCCATACCCAGCAGCGTAACGGCGAATGCAAGATACTTTTTCATAAAGGTTCCCACTGAATAACGCATTATGGGATGAATTGACCCTGGATTGGAAACCGAGAAAGTGATCGAGCCAGCAATATTCTTTGCCGGCATCCTTTATTTTCTCTTTATTGAGGTTGTATTGATAACCACAGCCCTGTGGCAGGGAAGGGGAACAGAACCTGTCCTGACCTTAGCTATCACCACTATCAG
	7.07	2.68	3.49	14.57	1E+06	STM1095					
	5.43	3.21	0.49	6.35	1E+06	IR STM1119 - STM1120	wraB	trp-repressor binding protein	-		TGTAGCGATTCGCTACGTCTATTTAAAGATATGCTCTCCTGTGAAGAGTGCAAATTTCAGCGCCATTTCTTTGATTTATAACAATAATTAATTTGGCGACCTTTGTTGCAAAATGATACATTTTTAAGCGCTTTGATTTTCCCAAATATAAGAATAACTTATTTATTTCTTATGGTTATTATTCTGCGTATTCGGCTTCCAATGTTGCAGAATATTTCGGTAAGCGGCCTACTACGACGTTTTTCACTATGCTTAATGTTACGCGGCGTTACTGATGATATCGTTCATACGCTGCGCGAGG
	2.81	5.09	0.80	5.56	1E+06	STM1120					
	5.74	4.54	2.14	8.31	1E+06	STM1186					
	5.68	3.84	2.94	13.36	1E+06	IR STM1186 - STM1187	STM1186	pseudogene; in-frame stop following codon 97; no start near coli start	+		CGGAAACCGCATCATTATTCCACTGCTAACCTTGTTATAGCAAGATGACTTTTACCATTTATCACCCGCTTACTCACAGTTTTTTCACCAGCGTGAGCCAATCGCTTTAATAACCAGCAAAACCGCAGTGAAAAATGTTCATCCACTGGCGTAGACGTCTCTATAAGCATAGAAAAATGTGTGGCGCGAATCTCACAGGCTATTTAGAATCGCCCCCCATGAAAACAGAAACGCCATCCGTAAAAATTGTTGCTATCGCCGCTGACGAAGCGGGGCAACGCATTGATAACTTTTTGCGCAC
	5.68	2.96	2.94	12.77	1E+06	STM1187					

	22.75	1.36	4.14	4.13	1E+06	IR STM1224 - STM1225	sifA	lysosomal glycoprotein (lgp)-containing structures; replication in macrophages	-		ATCGACCCTTTTTATCTCAACTGCGGG CGCATCGGATGTAATATAATTTTTAAAA GAGACTGGCAATCAGTATAAAACCTGA GAGCTTCGCGTATAAACGCATTACTGT CTGTGATAGCGTCGCTACAGGTAAAAA TAAAAGAAGGACTACCGCGGATGATGT TGTAGATTTGCAATACTGGCGGCAACT TCTTTCATGCGTTTTTTATGCCGAAGGC ATGAAGTTTACCCTTGAATAAACTTCAT GCCTGGATGCGTGTGGATTTGTTAGCG TTGCGCAATTAATCGCTTATATCACTCA
	18.59	1.38	3.56	2.15	1E+06	STM1225					
	11.41	3.53	2.69	5.70	1E+06	STM1262					
	12.43	1.43	2.63	3.49	1E+06	IR STM1262 - STM1263	STM 1262	hypothetical tRNA	+		GGCCGCGTAATTTTTCTTCCGCCATTA GCTCAACCGGATAGAGCATAGAGCTTC TACCTCTAAGGTTCGGGGTTCAATTCC TCGATGGCGGACCAGTTGATATCAAAA AAGGCCACCTGCGCGGTGGCCGCTGA GTTTCTGTTGAAATAAATGCAATGTTAT AATATAACAATCATCTTTCTAAGAAAGA TGAGGGTAACGTTTTGGTGATTCATTTA AAAAAACTGACAATGCTTCTGGGAATG CTGTTGGTAAATAGTCCTGCCTTCGCG CATGGTCATCATGCTCATGGCGCGCCG AT
	11.54	1.35	2.48	3.35	1E+06	STM1263					
	13.02	1.20	2.58	5.66	1E+06	STM1270	yeaS	paral putative transport protein	+		
	15.43	1.23	2.41	5.51	1E+06	IR STM1270 - STM1271					TTCTGGCGCTTTTGTAACCCACTATATT GGTACCAAAAAGAAACTGGCAAAAGTG GGCAATTCTTTGATTGGCCTTCTTTTCG TCGGATTTGCCGCCCGGCTGGCAACG CTCCAGTCTTAACCACCTGGACCCGTC GTCAACGGCGGGTCATTGCTCTCCTTT CGGTTTTATTGCGTGGAAAACAGCAAA ATAGTAACCAATAAATGGTATTTAAAAT ACTGTTTTTGGAGCGTAACCTTTTTACG ACAGCGATGAGATTATCGCTGAGTAAC CTGCGTGAAGAGGGAAGCAAATGCGG CA

	13.99	2.43	2.21	7.19	1E+06	STM1271					
	5.67	2.83	1.08	7.64	1E+06	IR STM1311 - STM1312	osmE	transcriptional activator of ntrL gene	+		CGCTGGATGATACCGGGCACGTGATTA ACTCCGGCTACCAGACCTGTGCGGAGT ACGACACTGACCCACAGGCGCCGAAG CAGTAACAACTGTACATTGCCTGAACAT TCAAGGAAACCGGCCTGCGAGCCGGT TTTTTTGTGCCTGCCATAACCTTATTTA TTATCGCGAATTATTTGCCCGAAATGTG AGGGGGGTCATAACGCCAGGTCAATG AGAGACAATTTAGTGGGTCAAGGAAAT ACCATCCGGTGGTCCGATCCCGTATAC TCATTTCAGCCACCTAAAAAAGTAAATC CGG
	3.10	2.03	2.19	3.50	1E+06	IR STM1360 - STM1361	ydiN	putative MFS family transport protein	-		TTATTGCATTGATAGCATTTCATTTGTTA GCCAGGAAATATAAAAATTGCTGCGAA TTTGTTGTTTAATACATATAACTCGTGA TGCTCATCGCAATTTTTCTGATAAGTGT GAAGATAATGAATAATAATTAACACGAA AATTACATTTTTTGTTTCCCGGTGATAA TGGCTAACGTTTTATTTTGCATAGCAAG GCAATAATATTGCAACTGGCACGCTAA CATTTATTGCGCGGTTGACGCTGCTTC AGCGTGATGTTGTGATTCAGCCCGACT TCGGTAACCGATGAACAGTGCGAG
	4.06	6.04	2.68	4.86	1E+06	STM1361					
	5.49	3.54	0.64	6.24	1E+06	STM1364	ydiK	putative permease	-		
	5.96	2.50	1.73	12.49	1E+06	IR STM1364 - STM1365					GCTGTACTATCCACAAACAGGCCACAA TCATGATGGCTAAAAACAGCACCGATA GCAGCACTTGCGCAATATCCCTGGGCT GACGAACATTTACCATAAATACTTTTCA CCTTTGTCTTTGCGCCAGAACGTTGGC GCGACGTGAACATGCAAACCACACCCT ATAATGATGAGCAATTTCAGCGGTTTTT AACAGGCCGATTCTGCATGTAATTCTG TTGGGCGCACAGGAAAAAAATGTGATA CAACAAATAACGCAACACGCAAACGAT TAAGCATCCCTTCCTGTGCGTAGACCG CT

	11.27	3.11	0.89	6.43	1E+06	IR STM1377 - STM1378	lpp	murein lipoprotein, links outer and inner membrane s	-		TGATCGATTTTAGCGTTGCTGGAGCAA CCAGCCAGCAGAGTAGAACCCAGGATT ACCGCGCCCAGTACCAGTTTAGTACGA TTCATTATTAATACCCTCTAGATTGAGT TAATCTCCATGTAGCGTTACAAGTATTA CACAAACTTTTTTATGTTGAGAATATTTT TTTGATGGGAATGCACTTATTTTTGATC GTTCGCTCAAAGAAGCATCGAAATGCA TGAAAGTCCCTAAAAAACCGAAAGAAA ACAGGGGGCTTCCATCGGATTCTTCTT AGATAATCCGCAATTAGATAGTAAAA
	12.11	2.11	5.46	4.68	1E+06	STM1389					
	14.05	3.53	5.48	6.58	1E+06	IR STM1389 - STM1390	orf319	putative inner membrane protein	-		CTTATGTCCGCCATCAAAGCGTACCGT GGCGCCAGTCAGACATCCGCTAATGCC GACTACGGGTTTGTTATTCATGATTCCC CCTTATTGAAAGTACGACGACTGACGC CAATGGCGCAAAATGTTATCTCACGCT GATTTAAAACTTACACAACTTTGTTTTTT TGTCTAAGTTTTCGCGGAGATTTTTTTT GACGTAATTAAATATCAATAAGATAGAA TGAGGGGAAGAAATCTATTTCAGCGCC TATAGTGTGATAACCTCCAGCGAAGCG ACCACGTTGCGCCACTGGGCAAGCTG
	14.85	3.17	5.44	8.13	1E+06	STM1390					
	8.78	2.81	2.05	9.37	2E+06	STM1437					
	4.15	1.85	4.61	5.34	2E+06	IR STM1437 - STM1438	ydhM	putative transcriptio nal repressor (TetR/AcrR family)	-		AAAACGACCCTTTAGGCACTTGGGCGG TTTTGAGCAACTCGCTAAGCCCCATGC CGGTAAAACCCCGTTGCATACAAAGCT GCTCGCCGGTGGCCAGCAGATGTTCG CGGGTATCGTGTTCGGTTTGCTTATTC ATAGCAGGCAGTATAGTAGACCAGTCG GTCTACTACAAGCAGAGTTGCCATAAT GTCAGTTAGCGTCTTCAATAGTCATAAG CGTCAAACGTTGAGGAGGGGATGTGG CCGAGCAGTTGGAGTTTTTTCCTGTAG CAAGCCCATGTCGCGGTATCTGCCAGT CTGAT
	7.00	3.17	3.39	4.75	2E+06	STM1463					

	9.41	3.20	4.26	6.11	2E+06	IR STM1463 - STM1464	add	adenosine deaminase	-		TCAAGGTGGCGGTGGATGTCAGTCAAA GGAAGCGTAATATCAATCATGGGCGCA CTCAATTTTTAATAAAAGTGCGCACCAT TATACTACAGATTGATAATGCTCTGGAA ATTTTGCAAAAACGGAGTCATTACGTTG CAACTTCGCGAGAGCGCGGGAGAAATT TTGTATCATTCTCTTTAACGCGCCCCCG GTCAGCTCACGGGGGCGTCTCTGTTAT CGCCTCTCAGGATAAAGGGTCAACCCC CCGCCTGTAGACAGTATCAGCGAACGG TGCGGTGGCAAAATCCATATCCGAGAT
	8.15	2.46	3.30	6.09	2E+06	STM1464					
	8.84	3.81	4.45	7.93	2E+06	STM1475					
	12.95	2.78	5.34	7.26	2E+06	IR STM1475 - STM1476	rstA	response regulator in two-component regulatory system with RstB (OmpR family)	-		TCACCACGCGGCTCAACAATGACATCA ATATCATGTTTCGCCAGATAAGCGGCA ATGAGAGAACCCACTTCAGCGTCGTCT TCAACAAATACAATGCGGTTCATATTAT AAATGGAGAATAGAAAACGCCAACATA CACCGCCTCTGTTTTCCCTTCCATAAAT CTTTTCTAAACGAGAGCGGTTCCGTTAT GCTACACGCTGTTGTTATTAGCGTGTTA AGGCAAGGTAATGGGACTCGTGATTAA AGCTGCCCTGGGGGCGCTGGTCGTCG TATTGATTGGTCTGCTGTCAAAAACGAA
	12.88	2.12	5.34	5.77	2E+06	STM1476					
	13.06	6.41	3.01	5.77	2E+06	IR STM1588 - STM1589	yncB	putative NADP-dependent oxidoreductase	-		CTTGCGTGATATTCTCATCTTTTACAAC AATACAGGTTTCTTTATGGCAACCGTTT TATCTCCGTCATTCCTTCATGTATCGAG ATTTTTGACCGGTTCAGGCCGCTGAGG GAGATAAGCTGCCCCACCGCGATCTGA ATGATGAATATAAGTAAAGCCGCAATTT TAAAATTTGCACATTTTTATGGCGACAT AATGCCGCCATTTTTTCTTTACGCATCG TCCGCTAAACGTATCACGACTTTGCCA AAGTTCTTCCCCGCCAGCAGCCCCATA AACGCTTCTGGCGCATTTTCCAGCC
	12.88	6.41	2.39	6.58	2E+06	STM1589					

	6.40	4.19	4.85	7.12	2E+06	IR STM1651 - STM1652	nifJ	putative pyruvate-flavodoxin oxidoreductase	+		ACGCAATGGCCCAGCGACAAAATGAATATGTGACAATAAAGGCATATAACAGGCGTAGAATATCGTAACCGAATGATATTGTATAATTTTTATTTTGTATAATACCCCCAAAAGCATTCGTATAAATTATATCTATTTCACTGCGAATTATTTCATTAATTATTGAATTAAACGGTAACATCTCTTTTTAGGTCTTTCCTGACAAGGCAGAAATAACGTTTTAACGTCAACTCGCTGATTATTTACGTGGAATACGCGTAATATTACGTCGCCCTCCCCTGTAGGTAGTCCCCGCAGAGTA
	4.08	3.17	4.01	5.20	2E+06	STM1652					
	2.87	2.35	8.22	8.30	2E+06	IR STM1748 - STM1749	ychE	putative integral membrane proteins of the MarC family	-		ATGTTCGTTAATGATCAAAACGCGCAGAAGATACGCCTTTTATTCGCATAGTTCACCTCTTATCTACGCCTAATTTCATCCATTCATCGCTGTTATTTATATGTACTCGTTATGCTAATCCACTCACTCTTCATGATAACGATTTCTTAACAATTTACATAAAAGGCTAAAATGGCCTGCTGAAAGGTGTCAGCTTTGCGTAATCTTGATTTAGATCACACAATCGCTACTCAGAAGTGAGTAATCTTGCTTACGCCACCTGGACGTAACGCGTTAGAGTTAAATGATACTAACGCAGAAG
	3.34	1.80	4.30	3.36	2E+06	IR STM1752 - STM1753	galU	glucose-1-phosphate uridylyltransferase	-		CCCAATCCCGCGACCGGGATAACGGCTTTTTTGACTTTCGAATTAAGGGCAGCCATTTAAAATTCTCCTGGACTGTTCATGTATTGAACGTGTTCATTAATCTGTATCGTGTTCCAGTATATCAGTACCAGAACAAGCCTCAGGTCCAAAAAGGACTTATATTGGTATAATTAAGACAAATACTTATAAATCTGCCGCAGATAGTAACACTCGTCGGGAAAGGCCGGTAAAGCAATTTCCGCTCACTCTTCCGTTTGGTCATTCCGCAGACAACATCAATCGCAGACGCCCTCCTGCGCCC
	3.37	3.21	4.25	6.30	2E+06	STM1753					
	19.52	7.93	7.59	11.87	2E+06	STM1785					

	20.40	9.07	9.65	17.70	2E+06	IR STM1785 - STM1786	STM1785	putative cytoplasmic protein	-		ACGTCCCGAAAAAAATGAATCAAATAAT CGGATAAGTCAAATCTGATGTTATTTTT CATGGGACGCCCTCTTTCAAACAGTCT CTTTTTTGCATTCCTTTAAAACCAGCAT CACTATTTTATATAAAAATCATCACGAA GTATGCTTCTTTTAACGATGACCTCAAA TCCTCCCCCCTTTTGCATCAACTTACGC ATCCCTGAAATGGCGAGAACAGGCTAA ATCTACCCGAGGTCACTCGCTAAAAAC CTCATCCTGGAACAAGCTCAACCGCCC TTCCCCGCTACGGCCCTTTCGCCGA
	11.00	2.99	0.32	6.05	2E+06	IR STM1794 - STM1795	STM1794	putative homologue of glutamic dehyrogenase	+		CCCGCCGACAGGACGACATAACATTGA TACATGTCGTTATCATAACGTTTACTTT TAGAGGTGCGTCATAATTATGACAAATA GCCACCTTGCACATATTTCGCATATTTA AGCAATTAATTGCATAATTAGCAATATA TCACCTCTTATAGCGGATAGTTAACCAC TTCCCATCCAAAATCATAACGAAAATCC AACTGCCTGCCATTTTTGATCTGAGTTA ATTGTTTAAAAAAGTGTTAAATTTATCG CTACATGGTGTGATCTACTATGTACCAC GGTCAATTAAAGAACATATTAC
	10.76	3.19	0.36	5.54	2E+06	STM1795					
	8.86	4.20	0.89	13.00	2E+06	STM1813					
	8.17	4.02	0.89	14.31	2E+06	IR STM1813 - STM1814	ycgL	putative cytoplasmic protein	-		CGAATCCTTTCATCAACGCTTCAGGCA CCCGCGAAAAATCGTCTTTTTTTTCGAC ATACAAATAGGTTTGATCGCGCTTGCTA CTTCTATAGATCACACAAAACATACTTT TACTCTGAATTAACGGGATGGTGACTT GCCTCAATATAATACTGACTATAACATG CCTTCTGGACTTCGGAATATCACTCCG TATCGGAGATGATAAATAGCAAATTGA GTAAGGCCAGGATGTCAAACACGCCAA TCGAGCTTAAAGGCAGTAGCTTCACCT TATCAGTGGTTCATTTGCATGAAGCGG
	7.85	3.58	0.82	13.13	2E+06	STM1814					
	5.50	8.38	4.89	4.63	2E+06	STM1839					

	5.50	9.75	4.99	5.51	2E+06	IR STM1839 - STM1840	STM1839	putative periplasmic or exported protein	-		CAATAACGCTTCGAGCAATTCTATCTGCTCGTTGGCACGGGAGCTTGCCCGGTTGACAAAGAACCAGAGCGCCAGCCCCACCACCAGAACCACCATTGATACTATTAAAGATGCAAGAGAAAACGCACCAGAGTTTAAAACGTCGTTCATTTCACCACCTCAATGTAGAGACGTCATTCTACCACTGCTACACGGGAAGGAAATCTCTGGTGTAAAACGTTTACCAGGGAATAAATTTATTGATGGCGCAAATACCGCTGAAAAATTGTACATCCTGATCGCACATGATATTAAACACCTG
	5.70	7.66	4.99	8.75	2E+06	STM1840					
	4.69	4.19	4.44	7.68	2E+06	IR STM1840 - STM1841	yobG	putative inner membrane protein	-		AATTGTACATCCTGATCGCACATGATATTAAACACCTGCGCCCACAGCAACAGGCATACTACCACCACGATGCCGAGAACGACCCATCGAAATTTTTTCACTCCACTCTCCGATCTTACATCTTATGTCGCTAAATTATCATGAGTTACTTAAACCAGGAGTAACTGTAGCGGCATTATATGTTTTTAGGAATGATTCACTTGTTTCAATCAATGTACACGCTACTCTTATTCTAACTAAAAAAGAAAAGAGGTAGTAATGCGTTTGATCATTCGCGCAATTGTATTGTTTGCCCTGGTGT
	3.83	2.95	3.54	4.78	2E+06	STM1841					
	12.66	3.22	3.87	6.92	2E+06	IR STM1855 - STM1856	sopE2	TypeIII-secreted protein effector: invasion-associated protein	-		AAACTACAAATGAAATGGATTGACGCATCTATTAGTGGTCAAAAAAACGCGCTACGAGAAATAATCAGTAACAATTGCAACACTATTCCAATCATAACGTAAACTATATGATACCAGGTGATTATTATTGCTTTTAGGTAACATATCTGTATGGCTGCTTTTAAGCAACAATACTCTAACACAACATATAACATTATAACTTACAATAGGTTAACAAATGGAATTACAGCTTATGCTTAACCACTTTTTCGAGCGCGTCAGAAAGGATGCAAATTTCAACGCATTTCTAATCGATCTGGAA
	11.89	3.22	3.87	7.20	2E+06	STM1856					

	19.06	3.74	0.57	7.84	2E+06	IR STM1866 - STM1867	STM1866	pseudogene	-		TGATTTAATAAGAGAAAACATATTATTA CCCTCATAGTAAGCAGTATTAAATAAGC CGGGATATATCTGATGTTCAATCAGTC CCTCATATAGGGTTAGCACCATAGCGA GTCGTTTTCACAAAAAACACAGACTGTT GAAACTTTATTTATCACTTTGACATTTG CAATACATGACACATGATTAGCTTCAGC CGCCATTATAGGGAAAGCTCCATTTCC ATACTCATTTACTCACTTCTCCCTGCGG AAAAAGAAATGCAGTATAGCCAGCGTG GTGCTTTTGCTGAAACCAGGCGCGA
	5.10	5.03	3.26	16.52	2E+06	STM1933					
	4.54	5.03	3.36	16.19	2E+06	IR STM1933 - STM1934	STM1933	putative ribose 5-phosphate isomerase	-		ATGTACGTCAGGTGATGGTCATTTTCG TCGCACATGCCGACGTTAAAAACGGGA AATCCCTTTTCATTGGCGACGGCGCTA AGTTCGTTATAAATGATGGCATTTTTGC TGGCCTGGCTATTTTCCATCATCAGTG CAATTTTCATCGTGTTTCTCCTGAATGC AGACGGTCGCGCCTGCGTAAATCATGA CGTTTTACCCACATTACACATTTGAGAA CACACATTCAAATTTAATAAAACCAGGT TTCATTAAATGAAAAGACGCTCACACAT TTTCTGTTCCCGCTGTAAATCCCCTG
	3.30	3.86	0.86	10.98	2E+06	STM1957					
	3.72	2.84	0.98	6.19	2E+06	IR STM1957 - STM1958	tnpA_2	transposase for IS200	-		TTAATATGCTGCCTACTGCCCTACGCTT CTCTCCATAGAACGCTTGTCTTCGGTAT TTGGGCGCGAAAACTATGTGATATTTA CAGTTCCATCGGGTGTGCGCTAAGCTC TTTTCGTCCCCCATTGGGACCCCCTTTT GATTTCTTGTTGAACTTTTGCAGTTGCC AGACCGCAAGATGTTTTAACAAATCAAA AGGGGTTTTAATAACTGGCTTAAAGCT GAAAGCTTTCCGGAACCCCCAGCCTAG CTGGGGGTTTTCCATAGACAATAAACG GGATGCGCAAAAGCCCACCCCGAACA
	5.77	1.84	4.86	5.12	2E+06	STM1966					

	6.40	3.52	5.94	5.51	2E+06	IR STM1966 - STM1967	yedF	putative transcriptional regulator	+		ATTCCACTGGATGCGCGCAATCACGGC TATACGGTGCTGGATATCCAACAGGAT GGCCCGACAATTCGTTATCTGATTCAA AAATAAGCGCATACTCCCGCTGTACGT TACGGCGGGAGACCTTTTACGGCATAA CCGGCAAAAATCTACAACGCATAAAAG AAATCAGACAAGGTCGTCTTGTGCGCC GTGGCATAAATCTATTATATAACGTATA CCGTTTTAATTCTGTCTGAGCCGATGAA AAATCCAGGGTTATTTTAATCAAAACAT AAAACAATTATTATTTTCCGTCTACGCC
	5.61	3.99	3.98	9.77	2E+06	IR STM2147 - STM2148	thiM	hydoxyethylthiazole kinase (THZ kinase)	-		TCAGACTTCCCTACGCTGGCATTATCC AGATCAGGTGGTACGGGTATTTCTCAG CCTTCACAAAGAAGGGCACCCCGAGTC GTCAAGCCCCACCGTGTTAAGCGGGG TTTCGCTATTAAGCATACTGTCTGTGCC AGACAATGTAAATTTACAGTCAGCGGC GGACGATAATTTCAGCGTTATCAGATA GTTCTCAAAACCTATTCGGTTCTGGCAA ACTTGCTGGCGGATATGTTGCTGCACG ACGCTTTCGTTTACACTTTTTACGAAAA GGGGCGTGAGATAACAAAATAGCGCTT GT
	8.35	4.88	0.85	5.87	2E+06	IR STM2159 - STM2160	yehU	paral putative sensor/kinase in regulatory system	-		AACTCGTACATACCCGCAAACCACACT TCAATTAAAAGCGCGTAACATACATTGA GTACGATTAACTTTCTTTGAACTGTTGC ATAAAAATATGAATTCGTGAATACGATC ACTTAAACGCCGCGCCGCAACCCGCTA CTTCGCGTTTTAATGCATAAAAAACAGG CAAAACTTCCTGGTTCCTAAAAGAGCG TCTAAAGTTAAACCGGGACCTCGCGAG CAAGGGTGAAACGATGGCGCTTTACAC AATTGGTGAAGTGGCTTTGCTTTGTGAT ATCAATCCTGTCACGTTGCGCGCGTG
	9.38	3.01	0.67	7.05	2E+06	STM2160					
	14.27	3.59	10.29	16.23	2E+06	STM2180					

	11.49	3.86	11.30	17.89	2E+06	IR STM2180 - STM2181	STM2180	putative transcriptional regulator, LysR family	+		CGCAACGCTATGCCAGCCAGGGGCAACTGGCGATTTTAAACTTGCCAAAAATTGAGCAAAAAGGCAGCGTAGGGATGTTCTGGCGTAAGAATGAGACGCCGTCTTTGGCCCTGAGTCGCTTTTTGTATTTTTTAGCCCAGGTTTAGCGCCGCCGACCAGGGGCATTGCCCGATGTTCCTGCTGTCTATACCCACTATGCTAAGAATTCATGATGTGATCGGTAGCACGTTTTAACGTTTAATTGTATGATGAATCCATCTCATCAAGGGCTTTAAACATGAGTAAGTCACTGAATATTATC
	3.94	3.73	0.47	5.79	2E+06	STM2226					
	5.04	2.26	0.41	4.33	2E+06	IR STM2226 - STM2227	yejK	nucleotide associated protein, present in spermidine nucleoids	-		GCGCTTGATAAGCTGGTGCAGGGCAATCTGGTTGATATCCAGACTCATGATAAACTCTCCTTTAAGACCGGGCGGTATTCAACCACCGCCTGCCGGAAGACGCAAGCAATCGCCCTGTCATTTCAGGCGTTATCCGTAACGCGAATGATTTAGGGGATAAAAATGCAGAAAAAAAACTGTTGCTACGGTAATATGTTGCCCTTTCATGAACAAACAGATTTTGATTTATGCCACAACTCTCCCGCTATAGTGATGAACATGTTGAACAACTGCTGAGCGAACTGCTCAGTGTACTGGAAAA
	4.73	2.38	0.36	3.82	2E+06	STM2227					
	6.87	2.44	5.79	5.78	2E+06	STM2280					
	13.11	3.72	5.26	12.44	2E+06	IR STM2280 - STM2281	STM2280	putative permease	-		CAAAAAAGATAATAAAACTGACTATGGTGATTGCCCAAAAATCTTTCGTCCATAATTTTTCTTTCATTCTTAACGACCCGCTCAGATGGCGCACGCAGGCAACGCTCAGCTCAACTGAACACCTATCAGGTGCGTCAAAATGTGATGTATTCGATAGAATCACAGTATAAACAAGTGCACTCTATTAGAAAAATTAATCGTTTTAATTATATTGATTAGGTTTTACTAATGACACTAACCCAAATCCACGCCCTGCTTGCCGTACTGGAGTACGGCGGATTTACCGAGGCCAGCAAACGGC
	11.78	4.41	5.49	12.44	2E+06	STM2281					

	16.05	5.97	5.10	11.78	2E+06	IR STM2330 - STM2331	lrhA	NADH dehydrogenase transcriptional repressor (LysR family)	-		AATACCAAATGCAACTGATCGGGATAT ATCAAAGAGAATTTGTCATACCTTTAGG CGTCTACAGATTTCTGCTAATGATGGA CGTGTAAATCTTGTAACAGCGTCAAATA GTTTACCGAGACGCACAGATACAAAAA CAATATATTGAACAATAGGTTATGTATA AAATCGCGTCATGATAATTAGCAGACA ACGCAGACTACGCCCCCGTTTCGGATC ATTATCTTAACCTAAAACCGCTATATTT ATAAGTATTATTACGAATAATCTTAACC TGGGATATGTTATACTAATCGGACCA
	3.75	2.85	0.51	3.73	2E+06	STM2387					
	5.29	2.67	0.65	3.05	2E+06	IR STM2387 - STM2388	sixA	phosphohistidine phosphatase	-		ACCCACAAGGGGTCAAGGGACGAACC GAATCACTGGCGGCATCGAGGGCTGC GTCGCCGTGACGCATGATAAAAACTTG CATATTGCACCGCTTTTGTTAACCAGTT TCACCAACACGCTTACCACATGCCCCT ATTGGCTGCGGCAAAAATGCGGTGGC CGGCATTGTGCCTTATCCATTCACTGA ATGAAACGCTGTTTTTTACCTCAATGGC GTAAGTATAGTCAATCCTTGATTATTAT TTCGCCACTAAGGAGGCATTCAGTGCG GATTCATATTCTCTTTGACCTCAATTTC CCT
	5.41	1.95	3.44	6.00	3E+06	STM2408					
	8.14	3.92	5.34	6.93	3E+06	IR STM2408 - STM2409	mntH	Nramp family, manganese/divalent cation transport prortein	-		GGGTACGGGTGATTACTTTGATAGTGT GAAACGATAGACCGATACGATGACGAC CTGTATCAGAACAGTTTGGCTTAACATT ACAAGATTAGCACACTGATATAACTTTT CATTTTCATATTCAGTACAGTAAAAGTG TATTACAGATCACTAATTTTGAATCTCG TCACAGGTCCTTATTATAGTGTGTGTTG GATCTCGTTTTCTTTACGGCTGTTGCAT AGAATGTGCACGAAAATTAAACCTGCC TCATATTTGGAGCAAATATGGACCGCG TCCTTCATTTTGTCCTGGCGCTTGC
	8.86	3.00	3.70	8.75	3E+06	STM2409					

	10.45	2.23	1.34	4.06	3E+06	IR STM2481 - STM2482	acrD	RND family, aminoglycoside/multidrug efflux pump	+		TTTCGTGCTGATACGTCGCCGCTTCCC GCTGAAGCCGCGCCCGAAATAAGATCC CGGCCAGCCTGATACGAGGTGTCGGG CACAAAAAAGGCGACTTTCGTTGAGTC GCCTTTTCTTATCCCCTATGGGAGCGC GGTGCCTTCCAGGCATTTATTTACGAA GCATGACTTCGATAAAATCTTTCCAGTT CCCCAGTTCACGTTCAATCATAATAGC CTCTCTTATTATTATGGGTATTCTACGT AGTTAGCGGTATAGAGAGAAGTTCATT TAACCGATTGTTGCGATATCCTCTGGTT AT
	4.94	5.33	3.12	6.24	3E+06	IR STM2525 - STM2526	yfgB	putative Fe-S-cluster redox enzyme	-		ATTTTTGTTTCTTTGTTAGGAACTACCG GGGTACTGCTTTCAGGTGTGACAATTT GTTCAGACATATGCTATTCCGGCCTCG TTATTACACGTTATGGCCCCTGGAGGG TTGAAAAAAGAAACGCCCCGGTAAGCT TACTGCTCGTCCGGGGGCGCTGCATT GTACAAATTCTGGCGTAAGGATGCCAC GTCTGCACGCGGCATTAGCAAAAATAA TATTTGAACCGATAATTTATCGCCAACG CATTTACAGCGTGAAAGACGAAGGAGA TTAACGGGTGCGCGGGCACACTTCGC CTTC
	5.95	5.20	2.67	6.90	3E+06	STM2526					
	9.22	2.69	1.21	5.94	3E+06	IR STM2555 - STM2556	glyA	serine hydroxymethyltransferase	-		ATTCTTCGATAACAGGTCTTGACAAAG GTTTTTACGCAAACGATTACCTATGCGT CAGATAAGGGTTTCCTGAACGAGAGTC TGACGAATTTCAACGGATTTCTTTTCAG CTTTGTGATGCAGATTTTTCACGTTGTT ACCTCCATAACGTAAAGCAGAGAAGAT CCATTTACAATGCAAGGGTATTTTTATA AGATGCATTTGATATACATCATTAGATT TTCACATAAAGGAAGCACGTATGCTTG ACGCACAAACCATCGCTACAGTAAAGG CCACCATTCCCCTGCTGGTTGAAACA
	8.94	2.69	1.33	6.15	3E+06	STM2556					

	2.71	2.57	0.72	2.90	3E+06	IR STM2583 - STM2584	lepA	GTP-binding elongation factor	-		TCTATACGATCTATAAACCTATAAACAC GGTTACAGTCAGTCCTGACTAAACAGC AGCCGGCCTACCGCAGTCACGTTCTTG CAGACAACGTGACTGCGGTAATCCATC CCACCGGATTGTCTTCAAATTCTCCATG TTGCTGAATCGGCTAACAGCTTCTTAAA CGATCGGTATTAGGCTAGGTTCTAAAT CTTGCCTGAATGAAAATAAATGTAATAA TGATAGCTTGGTATTGACATATAGATTG AAAAAGCGCATGAAAATAGGATTCCAA CCAGCCATATTGCAATATGCATATAC
	2.68	2.44	0.60	2.97	3E+06	STM2584					
	4.64	4.54	0.35	9.55	3E+06	IR STM2620 - STM2621	STM2620	Gifsy-1 prophage	-		GAGTTGTAATTCGTGCGCCATGGTATT CTCCGTGGCGCATAATTGTCAGGTTAC TGGTTGTTCAGGCCAGTGCGATAATTA TGATTGCGTGCTTATTGTTAAGTCAATT ATTAGAGCCCATCTCTCTGACAACTTCC ATAATGGTATCCTTAGACCAGTACAAAT CATCTCTTGATAGCTTTCGGTTTGTTAC AGAACCGTTTTGGGTTGATAGGATATA GCGATCACCTGATGCAAGTCTAAAAGA CAAAAGTTCTATTCCCTTAGCATCAGTT ATAGTCACCCGCAAATTATCAGCAAG
	15.54	2.48	3.54	0.65	3E+06	STM2640					
	19.02	2.48	2.07	4.04	3E+06	IR STM2640 - STM2641	rpoE	sigma E (sigma 24) factor of RNA polymerase , response to periplasmic stress	-		ACGCACTATCTGTACAGAAATGCCCAT TTCGTCGTTTGCAGAGTAACCTAACAG CATCTTTATTTCACTACAAAATCCGACG CTAACACCCTGCCCTATAAAATATTTTT TGCCGTTTATCTCTCGCCGTATTTTTAT TTTATGTTTAATAAGCACAACACCAGCG AAATCATAACGTGCTTTTTAGCGCCATA TAGTGCTAATCTGCCGCAACCATGTTTA GTAAATTAAACAAGAACCATGATGACAA CTCCTGAACTGTCCTGTGATGTGTTAAT TATCGGCAGCGGCGCGGCCGGAC
	24.48	3.33	2.75	0.49	3E+06	STM2641					
	2.86	3.90	1.67	13.85	3E+06	STM2659					

	9.64	5.65	5.87	7.55	3E+06	IR STM2659 - STM2660	rrsG	16S rRNA	-		AACGAAGCTTTTCTGACCCGGCGGCCT GTATGCCGTTGTTCCGTGTCAGTGGTG GCGCATTATAGGGAGTTATTAGAGCCT GACAAGACCTAAATGCAAAAAAAAGCT CAACCGTTCACTTTTCAAACAACATTTG AACCAAAAGCCTATTTTCGCCTGGTTTT TAAACAAAAACGAGCCCGTCAGGGCCC GTTTTATTCAAATTTGTGACTTACTGCA CTGCCACAATACGATCATCATTGGCTT CAAGGCGAATCACTTTGCCAGGAACCA GTTCACCAGACAGGATTTGCTGCGCCA G
	19.87	1.84	2.99	2.17	3E+06	STM2662					
	4.23	6.25	3.58	7.92	3E+06	IR STM2662 - STM2663	rluD	pseudouridine synthase (pseudouridines 1911, 1915, 1917 in 23S RNA)	-		TTGACCAACACGCGCTGATTCAAAATC CATTCTTTTATACGCGAACGTGAATAAT CCGGGAACATTTCGGCCAAAGCCTGAT CTAAGCGTTGACCGAGTTGGTTTTCGG AGACCGTTGCGGTGAGTTGTACTCGTT GTGCCATATACAGCTTCTTCGTTTAACG TTGGGTTTTACGGCTTTGCCGTTTAATA TAGTGTGCTATTGTAGCTGGTCTTAACC GGGAGCAGGAACAGAGAATCTCCCGT AAAACATTTTGAGGAAAGTCAAAACGTC ATGACGCGCATGAAATATCTGGTGGCA
	4.14	3.10	1.03	4.32	3E+06	STM2663					
	7.50	1.89	3.23	2.75	3E+06	STM2801					
	12.46	5.53	4.30	4.62	3E+06	IR STM2801 - STM2802	ygaC	putative cytoplasmic protein	-		ACGGTAAACCCTGCCTTTTCCAGTACC CGCGCCACCTCGTCAGGTCGTAAATAC ATATTTTATCCTCATTCTCTTGTACTGC GGGCTTACCTTACCCGATAGCGCGTTA TCAACGCTTTCAGAAAAGTCCAGAAAC GCATGATATCGCCGTAACAAGCCTCAG CAGGTAAAAATATGAACTACACTGAAA GCTACATCGAAATCAATGGAGGATCAT ATGCTTAACAAACCGAACCGAAACGAC GTCGATGATGGTGTTCAGGATATTCAG AATGATGTCAATCGATTAGCCGACAGT CTG
	13.01	4.82	4.47	4.62	3E+06	STM2802					

	4.25	6.94	0.48	11.09	3E+06	IR STM2808 - STM2809	nrdF	ribonucleoside-diphosphate reductase 2, beta subunit	+		TCCCATGCCTTTATTTCAAGCAATAGGG AGTCAAATCGCGCAAATATTACAACATG TCCTACACTCAATACGAGTGACATTATT CACCTGGATTCCCCCAATTCAGGTGGA TTTTTGCTGGTTGTTCCAAAAAATATCT CTTCCTCCCCATTCGCGTTCAGCCCTT ATATCATGGGAAATCACAGCCGATAGC ACCTCGCAATATTCATGCCAGAAGCAA ATTCAGGGTTGTCTCAGATTCTGAGTAT GTTAGGGTAGAAAAAGGTAACTATTTCT ATCAGGTAACATATCGACATAAGTA
	9.87	4.43	3.25	7.89	3E+06	IR STM2874 - STM2875	prgH	cell invasion protein	-		TGTATAATGCGTCTCAACACATATTAAA AGAACCATCATCCCCATTGGGGCTTAA ACTACTGTAGATAAATTACCCAAATTTG GGTTCTTTTGGTGTAACAATCAGACCAT TGCCAACACACGCTAATAAAGAGCATT TACAACTCAGATTTTTTCAGTAGGATAC CAGTAAGGAACATTAAAATAACATCAAC AAAGGGATAATATGGAAAATGTAACCTT TGTAAGTAATAGTCATCAGCGTCCTGC CGCAGATAACTTACAGAAATTAAAATCA CTTTTGACAAATACCCGGCAGCAA
	9.87	4.47	3.25	8.16	3E+06	STM2875					
	3.68	4.26	0.55	5.31	3E+06	IR STM2903 - STM2904	STM2903	putative cytoplasmic protein	-		GGTTGTGTCCCTATTACGCGGGTAGGA TCAATCAAGCAGTTACGGCAAAAAAGA GAATCATGGATATATTTAGCAAACTCCC TGATGATACGTAATCAGTGAGATTAAAA TAATGCAATCGCGATAAACCGAAGTTA ATCCCCTGTTTAAAGACAGTGAGCGAC CTTCTTGCCATGCCTGGACTATATCAG CCTCATATGTACGCCTTGAAAGCGTAC AGATATGTATTATAATTGTACATATTGTT CATAAACAGGAGGATGAAAACCATGCC TCAGATAGCTATAGAATCTAACGAAAG
	3.81	2.82	0.55	5.19	3E+06	STM2904					
	4.30	2.81	0.47	5.50	3E+06	STM2954					

	3.43	3.95	0.42	4.50	3E+06	IR STM2954 - STM2954.1 n	mazG	putative pyrophosphatase	-		ACTTCATAGGTTTCTTCCAGCGTATAAG GCGCGATGCTGGCGAAGGTCTGCTCTT TATCCCACGGGCAGCCGTTTTCCGGGT CGCGCAGGCGCTGCATGAGGGTGAGA AGACGGTCAATTTGATGGTTAGTTGTC ATGGTTTTTAATCGGTTGTAAATACCAG CGACAATTGTAACGTATTATTCTTAACC ATTCACGCACAGAGACACTACGACAAC GCCTATATAATAAAATATATTGTTAACA GGTGTTGAATGCTACCTTTCCCGTATAA CTTTAAAATTATTAATCGATACACAAC
	10.45	4.17	2.04	7.90	3E+06	IR STM3016 - STM3017	araE	MFS family, L-arabinose: proton symport protein (low-affinity transporter)	-		AATGGCTACGCTATAGCGATATGTGAT GGATATTACACTTTTTAAATTTAACGCC GTTGCCGGGTATTTTTTTAAACCACCAA TATTTCAATGAATTAAAGCATTGATCAT AGCTATTATTTAACAATATATGGATTAA GTTAAACCCACAATATGGACTATGCTAA TGAGATCATAAAAAAACCCTGTACGAG GACAGGGCTTTATCAGTTTTTTCGGCC AAAGCGTCGATTTTCCCAGAAACGCAT TTGTCAGTAGCGGATTAACGCGCCAGC CAACCGCCATCTACCGCTATGGTATA
	9.65	4.43	2.52	14.23	3E+06	STM3017					
	2.67	2.05	2.00	6.06	3E+06	STM3023					
	3.43	1.93	2.11	6.54	3E+06	IR STM3023 - STM3024	yohL	putative cytoplasmic protein	-		TGTAACACGGCCGCGCATTCATGCGGT TCATCCAGCATTTTTTTTAGCGCTATCA CCTGTCCCTGAATCTTGCTGGTTCTGG CTTTAAGCTTTTGTTTGTCCCGGATGGT ATGTGACATTACAACACCTCACTAACAT TAACGAATACAAATTATAGCATTACGAT GCTACTGGGGGGTAGTATTCTATACTG GGGGGGAGTAGAATGACGCCCACATA AAACAACTAAGAATCATTCTCATGGGTG AATTTTCGACACTTCTTCAGCAAGGAAA CGGCTGGTTCTTCATTCCCAGCGCCA
	3.14	1.93	2.06	7.47	3E+06	STM3024					
	3.46	3.76	1.45	6.82	3E+06	STM3059					

	3.46	4.12	1.38	6.74	3E+06	IR STM3059.S - STM3060	ygfB	putative cytoplasmic protein	-		ATGAGCTGTCGTTGTTGCCGCCGCAAA TCATCCCGCTGATTAAACCATGCATTTC AGCCGGGGTCAGACCGGCCCCTTGTT GATTCAAAAACCGGTTCATTTCGTTGTA ACCAGGCATTTCGTTCTGTATAGACATA AGCATTCGTCATCAAAGGGAGGATATT CATGATATGCTACCACTTTGGACCCTG GTGAACCAGAAAAGGGCTTGTATCTTC ACACCAGGGTAGCTATAGTGTCGCCCC TTCGCGGACCCTGGGTCTGGAGACGA AGGCAGCGCAGTCAATCAGCAGGAAG GTGG
	8.64	3.59	3.25	2.57	3E+06	STM3060					
	10.29	5.01	3.53	9.98	3E+06	IR STM3062 - STM3063	serA	D-3-phosphoglycerate dehydrogenase	-		CTTTTTTGCCATCTGATGTTGTGTGTGG ATTTGCATCCGTCCTTCAACATATCAAA AAAAATTATCACGGCAATATGAACGTTT GCGCCAGCGTCGTGAAGGAATCGCAT ACAGCGGGAAATAGCAGATGAAAATAC CGGGAATAACTTTTTCTTTGGAGGGAT CGGCAGGGCAAACGATTAAACGTGATA CATGTCACCAAATTTGCCCTGACCGAA TTTTTTACGCGGCAGGAAATACGCCTG GCGGGATCATTTTACGATGGTTTTCAC CCCGTCCGGCGTGCCGATCAGTGCGA CAT
	10.25	4.50	3.68	9.08	3E+06	STM3063					
	8.70	6.90	4.94	2.66	3E+06	STM3083	STM3083	putative Mannitol dehydrogenase	-		
	6.87	6.27	5.83	3.36	3E+06	IR STM3083 - STM3084.S					TGAGATCGTTATAAACAGCCTGATGAC CACGGTGAAAGGCGCCAAATCCAATAT GTACGATGTTGGCTTCCATTCCCTGAC GTGAATAAGTCGTTTTGAATTGGTGCCT TGCGGCGTCTAACTGGCGAGCTATGGT GTCCATGAATTTTTCCCACTCCTGTTTT GTTTACCAATTCTGCTTAAACACCATAC CAAAATCCGTGAATATGATCACACTCAT GGCACCAGATTCTTTACCATGGTATGC TGACTAATAGCCAATGAATAAAAATAAT TTATTTATCAATTAGTTATAAAAAGC

	8.91	3.97	0.22	11.50	3E+06	IR STM3168 - STM3169	ygiR	putative Fe-S oxidoreductase family 2	-		TGTTTGAAATTGGTCTTATGAATATCTT CAAATTGGTATGCAATTAATTATACCCA CGTCTAAAAACGCAGTATCGTCATAAC AACAAAAAGTAAAAAAACATCACATTAT CAGTAATATATAAAAAAACTTCGCTGAA TTGCTCACGACACTGTTTTTACCATGAC TTTCTTCTGTGAACCAGATCTCTTTCTT TGGTCTATTGATTAAATTAAATTGGCTG ACAGAATTCAGGGGATAAAGAACACCA TCACCACGCCTTTCCCCAACGCAACAC CTTACGTATCAGCAGGTTATTAAT
	8.70	5.18	1.38	13.67	3E+06	STM3169					
	4.81	2.12	0.39	3.00	3E+06	IR STM3195 - STM3196	ribB	3,4 dihydroxy-2-butanone-4-phosphate synthase	-		TCCGGACTTTAACCGTCGGCCCCGGAA TTACACCGGATCTGCTGACCTTTTCGC TATGGCAAAAAGCGCTCGCGGGCTTTC AACCTGCTCTCCGCGTTCCGTCACGGC GCGCCGTGATGAGAAATGCGTTAAACA TCGCTGATTTACCGCCGGTGGGGAATT TCGCCCCGCCCTGAGAATAAGCGGGTT AACTATAACGCTATTGATTACCTTCATC AACGCCTTTACTCCGTATGACGTCACA CAATTCTGGTTTATGGCGTCCACATATC GCACTACAATAAGAGCTAACACTTACC AG
	4.57	2.33	0.38	3.20	3E+06	STM3196					
	4.31	3.54	1.26	4.72	3E+06	STM3202					
	4.70	3.24	1.03	5.13	3E+06	IR STM3202 - STM3203	ygiF	putative cytoplasmic protein	-		GTTATCAGGCGTTTCGAAGTAGATATTC AGCAACTGGCTGGGCGCATGATGCTC GCCGCCGAGCGTATGAAGATGATTTCG CAGCGCATCTACGGCGTCGTGATTGAC GATAAACTTTAATTCGATTTCCTGAGCC ATGGCCTTGTACTTATGGGTTATGTCAC ATCTGGGAAGATTCTTGGCGAACTTAC CCGCATTATTTTTGTCAGTAGATAGTAT TTTGCGCCAAATTGCCATGCAACGAGC AATTTGACGGGCGTAAAAGTTTGACGT AGCGGCAAAGGCGACACAGATGATTCC G
	4.20	4.68	1.34	5.12	3E+06	STM3203					
	2.91	2.54	2.85	2.95	3E+06	STM3214					

	4.36	2.62	4.77	2.91	3E+06	IR STM3214 - STM3215	yqjH	putative transporter	-		CCCGCAGAACGATCAGCTCGCGAAAAC GCAGCTCATTACGAACACGCTGTGGGT AGCGTACGGATGATGTCGTCATTTTTT GCCTTCGTGAAGTAATACGATATATCTA AATTAAAGTTTTAAATGATAATGATTGTT AATCAGTAAAAATGCAACTGTTTTTTGA TAGTGTTCTGGCAACACATCGCTAATC ACAACTTCAAAATAAAACGTTATAAATT AATAGATTATATCAACAATCGCTTTTAT CCTTGCTAAAAACCATCATTTAGATATA AATTAGATATATCTAAATAAGCAG
	3.38	1.90	3.56	2.09	3E+06	STM3215					
	16.37	5.99	0.24	12.63	3E+06	STM3245					
	12.29	5.70	0.27	9.88	3E+06	IR STM3245 - STM3246	tdcA	transcriptional activator of tdc operon (LysR family)	-		AAAATAGGCCTCAACATCGCTAATGATT TTACTGACGGCGGGTTGGGTTAACCCT AACGATTTTGCGGCAGAACCGATAGAA CCACTTCTAATGACTTCCTGAAAGACCA CCAAATGCTGTGTTTTAGGGAGAACAA GAGTATTCATATCTACCGCTCTGAAATA ACATTGTGAACGGCAGGAAGTGTAGCA AATTAAATCTTAAAGGTTATGTGCGACC ACTCACAAATTAACTTACCACAATTTTT ACATGGTTTTTATTAAATAAAGAAAACC TGATATTTCAATAGGTTACAAAAAT
	2.46	4.21	0.82	4.51	3E+06	STM3297					
	2.33	5.69	1.36	8.16	3E+06	IR STM3297 - STM3298	ftsJ	23S rRNA methyltransferase	-		CAAGTTTAAACCAGGCACGGGAGCGTA GCCCCTTTTTCTGCGCCTGTTGAACAT ATTTATCGCTAAAGTGTTCCTGAAGCCA GCGGCTTGAGCTGGCAGAACGCTTTTT ACCTGTCATTTAACTTTCCCGTCGGGG CAGTTCATCGTAGCCAATGGCGTAAAT TTCTACACGCCTATTTGGCGATATAAG GGAGATGGCGGTAGAATGACCCGTTTT CAATCCCAACGTAAGCAAAAATATACG ATGAATCTGAGTACTAAACAAAAACAGC ACCTAAAAGGTCTGGCACATCCGCTCA AG
	2.78	5.49	1.44	9.14	3E+06	STM3298					

	8.69	3.03	0.58	9.26	4E+06	IR STM3342 - STM3343	sspA	stringent starvation protein A, regulator of transcriptio n	-		GACCAGAAAACAGCGTCATTACCGAAC GTTTGTTGGCAGCGACAGCCATGAAAA CCTCCAGGTATATTCAGAATTTTTACTG CTACCAGCCACAATGTGACCAGCCAGA TGTTATGTCACCCAGGGCGAAAAAAGC CATCATTGCTCAGAAACGAGACAAAAA ATGAACATTCCCCGCTATTTGGGCAGA AAATTGGATGATAGTTTACCAGATTTTG TGACCTTTGTGGTGAGTCGATTCTGGA AATGAGGAAAAAGAGATATTCCTGGTC TGAAATGCTCGCCCCACCTGAGATATT GT
	7.68	2.23	2.54	7.89	4E+06	STM3343					
	2.34	1.09	10.63	3.05	4E+06	STM3356					
	3.75	1.53	6.02	2.87	4E+06	IR STM3356 - STM3357	STM 3356	putative cation transporter	-		CATATTTATAATTATCCAATCAATGATAT ATGATATTGTATCCAATGTTGGCAGGG AGAAATTATTCCCATACAAAAACTAAGT CAAATCGTTTCTCAGGAAAGATGCAGG AGTGGGATCTACATCAAGATCGTGGTT AGATCGTTACTGGACGTGATTAATAGA ATTGAAGAATTGGTTGAAGCGCCTGCG ATGCTCACGCAGGCGAAAAGATCAGGC AGAAGGGTCACCAACATAGCGGGTCA GCATATTCTCCATTGAGCGAATAATGTG TTCGCGCATGCGCTGGCGTGCCAATGT T
	4.71	2.01	3.72	1.67	4E+06	STM3357					
	5.39	3.55	0.98	5.58	4E+06	STM3378					
	4.65	3.71	2.07	8.91	4E+06	IR STM3378 - STM3379	STM 3378	putative inner membrane protein	+		TAGCCCTTTTAGCGTTGCGTTACCGGA AGTTTCGCCAGTGGTGGCGCTAGTTTG GTGAACTGTGCGGTCGATTGCAAAACG CAAAACAGGTAATGTCCTTTTTATGTTT CGGGTTGATTATCTTCCCTGATAAGAC CAGTATTTAGCTGCCAATTGCGACGAA ATAGTTATAATGTGCGACTTTACATTGC CCAACGGCGATTTTCGTTCGCAGAAAG GGTGACAATCGAGCAATGAAGGTATAT TTTGTTTTTTGCCCGAAAATGGCAGAAG ATAGCCACACAATGACTGGCAAATCAT G
	8.32	6.32	2.17	10.71	4E+06	STM3405					

	7.92	4.90	2.30	8.48	4E+06	IR STM3405 - STM3406	smf	putative protein involved in DNA uptake	-		GTTCAGCTTGCCGCGCGGTAAGACCA GCCTCCTGAAGGTGCGTGCGATTTATC TGAGGCTGGCGAATAAGCGAGTTCGC CATGTTCAACATCGCCTCGCCATAAAG GTCGCCGACGTACATTAAACGTAACCA AATTTCGGTACGGGCCATCCTTTCCCT CCCCTGCCACAAGCAGTCTGAACAATC TTTGCGATTGGTCACTGATGCTGTCAAT CAGGTGGGGATTTGTCTAGAATAGAGG TAATAATCTTTTCAACTCCTGAACACAA CTCTGGATAATTATGTCAGTTTTGCAAG TGT
	13.47	1.74	3.60	2.98	4E+06	IR STM3453 - STM3454	fkpA	FKBP-type peptidyl-prolyl cis-trans isomerase (rotamase)	-		GATTTCATCCATATCTCCAGGGCCGGG GCATCTCGCCCCATGTTAACTTACGTA AGAAGCGTACTATAAATCGTTGCAGAA CAAATCAACATACGAACACGCCCTATTA TCACTTCTTTTCAGACTCTTTTTGTTTAA ATTAGTTTCGTAGTGCGCGTAATGGTT GCTGTGAAAGCCGGTAAAGTTAAGTAG AATCCGCCGACGGAGACAACATAAAGA GGTACATCATGCAGGATATCACGATGG AAGCTCGTCTGGCTGAACTGGAAAGCC GTCTGGCGTTCCAGGAGATTACCATAG A
	12.79	2.04	3.73	3.72	4E+06	STM3454					
	14.28	4.61	0.55	10.24	4E+06	STM3487					
	10.28	7.90	2.02	12.47	4E+06	IR STM3487 - STM3488	aroK	shikimate kinase I	-		AAAGATATTGCGTTTCTCTGCCATTTTT TCGGTACTACTAAGACTATTCGTTAATG GTAAACCCGCTTCACAGACACCCAGCG CAGCAGGACATGAACTGAAACCTCATA AGATATTGCGAGAGTCAGACTGAAAAT TATCTCAATACTCAAGCGGGTTTGGCA ACTGAATAAATCACCAAGCCTGATTGTT GCAAAACCCGAGTTAGCGTTGCCGAAT GGCGACCAGAACAACATATCCGGCCTA CAAATTGCTCTACTTTCAAACAATTGTG CGCAATCCGCAGAACCAATACGTCTGC

	11.79	2.63	1.44	3.45	4E+06	IR STM3494.S - STM3495	yrfE	putative NTP pyrophosphohydrolase	-		CACGCGACGCACGCCGTTGCTGAACT CCAGATCCACGCTTTCTACGTTAAACA GTCGGGATTGTGCGACGGTTTCCACTT TCAGAATGGTGGGTTTTTGTAATGATTT GCTCATTGTGAGAATCTTTGCAGTGTAA TCTGTGGTCATTGTGCGACATACCGCA CGGTTTCGGCAATGCGAATTGCCGTTT ATTTACATTTATGTAACGTAATAAAAATT AATTCTTATTTCAAATTAAAAGTCAATAG GTTGAAATAACTCCAGGAATTTGCTGAT ATTCCGTTTTTGGTGGTATTGCTAT
	10.33	4.08	0.35	3.90	4E+06	STM3495					
	19.41	3.10	2.01	7.35	4E+06	IR STM3504 - STM3505	yhgF	paral putative RNase R	+		TTAAACATTAAAAACGGTGAATATTTGC ACATTAGAGGTATTTGCAAAAAGACAAA TAAATGTTGAGCCATATCAACATCGGC GCAAATTATCGCTTATTTGTACATTCCG TCACATTTTAATCGTTGAAGATAGAAAC CATTCTCATTATCATTGTGTTGTTGATT ATTTACTCTTTCCTTCGTTGGCTAAACA TCGGGTCTCCTGCCGCCCCCCTGAGC GCCGCATGAGGTATACATCCAGTTAGT AAGAAACAAGTAGGTCGTATGCAATTC ACTCCTGACACTGCGTGGAAAATCAC
	14.38	3.01	2.01	6.02	4E+06	STM3505					
	8.26	3.35	6.09	4.90	4E+06	STM3511					
	9.21	2.28	8.65	5.12	4E+06	IR STM3511 - STM3512	yhgI	putative Thioredoxin-like proteins and domain	+		TGGTTGACGTCACGCTGAAAGAAGGGA TCGAGAAACAGTTGCTGAATGAATTCC CGGAACTGAAAGGGGTTCGCGATCTGA CCGAACACCAGCGCGGCGAGCACTCA TACTACTAAGATTTTCCCCGCATCCATG CCCGATGGCGCTTGCGCCTGTCGGGC CTTGTCAGCCCCACCGTAGGCCGAATA AGGCGTCTACGCCGCCATCCGGCGCT ATCAACCACATCTCATAACAATGGCCCT TCTTCTTTCGCCGATAACATGACCTGTG TCTCATAATTTAAATTTTGCCTGCCAGG GTC
	5.59	2.28	1.83	3.95	4E+06	STM3559					

	10.95	2.11	2.86	7.17	4E+06	IR STM3559 - STM3560	yhhV	putative cytoplasmic protein	-		CCCACGACGCGTGATGGTAACAGGCC CCCCCGTCACCGCACTTTCCAGGACTT CGGCCAGATTTTGCCGCGCTTCGCTAT AGTTAACCGTACGCATAAACATCTCCC CAGTTGTACATGTTTATTGTACAACAAA CATGTACAAAAAAAGAGCCATCAGGCT CTTTTGAAAAATTTTACCGCTTGCCGTT ACCGGGGGCGGCGCACGCGCTTCCCC CCTGGCACAGTCTAACCGCCCAGATAG GCGCTGCGCACCGCTTCGTTCGCCAG CAGTGCATCACCGGTATCGGATAGCAC CACGT
	10.33	2.11	3.06	7.27	4E+06	STM3560					
	7.59	2.00	1.08	7.04	4E+06	IR STM3590 - STM3591	uspB	universal stress protein B, involved in stationary-phase resistance to ethanol	-		AGACAATCAGTGAAAGAGTACTACGAA AGCCGTCCATATTAGCGCTCCGCATTC GAACGGCTCTTATACACATTGTAGGAG ATCAGTTAATTTTTTTACCAGAAGGTTA ATCACTATCAATGCAATTCCCTAGAAAT TTTGTTTAACTAACTGGCAAGCAAGGC AGATTGACGGATTATCCTGGTCGCTAT AATGTAAGGATAGTTATGGTAAACGGC TGAGCTAGCCCCGCGCATAGAGTTCGC AGGACGCGGGTGACGCGGCGGCATAA GAAACGCCAGTAGCTCAATGGTCATCG ACA
	5.44	1.39	2.01	5.66	4E+06	STM3591					
	5.41	2.58	2.89	4.25	4E+06	IR STM3630 - STM3631	dppA	ABC superfamily (peri_perm), dipeptide transport protein	-		TTCAGAAGGGTATTTTCAGCAGGGAAA TTTGTGCTATGGCCAGAAAGGCAGAGT TATTCACTTAATATTTTGCAACAGTTAG TGATTAACAATTAGACATTAATTGAAAA ATTTCTTTCGATATGTTGATTATCTGAG CGATTAATACCACTAACGCTAAAACGC ACAGGCGAAAATGCTGAGGTTATCCAT AAGCCGTGTGCAAAAAAGAGTTATACG GACGTTGAAAAACACCATCGAATATGT CACAAAATTGTAAATAAGTAGGCCGTC GTGCGGCCTACCGCGATCACAAAAACT A

	12.80	2.93	1.08	10.12	4E+06	IR STM3684 - STM3685	yibF	putative glutathione S-transferase	-		CATTAATAAATTCGAAGGTAATACCCTT TTCGAGCAGCAGAACAGAGATTTTGCG CACAAAAGGGCTGGTGTAGCTACCGAT GAGTTTCATGCCGTGTCCTTTTTGCCAA CCAGTAAAAATCATAGTATGGCTCAAAT AAGACGAAAAGAGACACAAAAGGAGGT TGCTGAATGACATAACGTGAGAGGACT CGCGACAAAATGTTTGTCGGATCGTAT TGACGTTACCCGGGCTTAAAATTTCTTG TGAAGAGGATCACAAAAATTCAACAAA GCACCAAAATAAAAATGTGAAATATCT
	3.23	3.46	4.44	3.72	4E+06	IR STM3793 - STM3794	STM 3793	putative sugar kinase, ribokinase family	-		TAAAATAACATTATCATGTTACTTCCGT ATCATTTGTGACTATGATCGCGATTAGA GGATCATTTTGCCATTTACTTCGTGAAC AATCCCTGGCGGAACATACGCGCACCA AATCATTTTTATTGTTACAATTTACTGAA AATTAACTATTTATTGTTATAAAACGCG AATAAACCCACTTTTATTTCCTGACAGC CGGACGTATAGTAGTGCCACACTGTAA TGTTCTCAGAAACACATAAATGTTACTG ATGGAACATAACAACATGATTTGCGGA GAGGGTGAATGGAGACCAAGCAA
	2.88	3.00	3.22	4.38	4E+06	STM3794					
	25.73	6.53	7.93	10.67	4E+06	IR STM3820 - STM3821	STM 3820	putative cytochrome c peroxidase	-		ACCCGGACAAACCTAAATAACATAACA GCCCAACGGTGATAACTGTTGTCGCAT AGAGGGTAATTTTTTTCATATCACTATC CTTATGGGGTATTGCGGCATGATTAATT AAATTTTATTTTTTTACTCATGAGGCCC GTCAATACTAAATACAAACCCATCATGG ATATTGATTGGTATCAATAATTACAATT GGCTAAACCTATAGATATGATAACCCC CGACTATCGTAAGATTTATTTTGCGATG TCCGTCACAGGGTTTATTCAGCAGCAA CAATGGATAAATCCTCTTTTCCGTC
	23.33	6.41	8.05	13.85	4E+06	STM3821					
	7.60	3.77	4.14	0.75	4E+06	STM3857					

	9.06	2.97	5.72	3.09	4E+06	IR STM3857 - STM3858	pstS	ABC superfamily (bind_prot), high-affinity phosphate transporter	-		CGATAAGGTCGCGGCGACAACAGTTG CGACAGTGGTACGCATAACTTTCATAAT GTCTCCTGCACGGTTTCGGTAAATCGT TGTTTGAGTTGCTACGATGAGCAAAATA GGACAAATTGATGACAGTTATATGTCTT GATTATGACGGTTTGATGACAATGGAA ATAAAAAAAGCTGGCCCGGGGAGACAC CAGACCAGCCTGCAGGGGGAGATGAA TTAGACTGTTTGCGCAACCGCAGACGG TTTCAACAGCGCGTACATCAGGCCGCA GACAATCGTGCCCAGGGCAATCGAGA GCAG
	9.06	2.15	5.89	3.60	4E+06	STM3858					
	2.26	6.29	0.46	10.23	4E+06	IR STM3899 - STM3900	yifB	putative magnesium chelatase, subunit ChlI	-		TGGCGTCATTTTCAGGTAAGAAACATC AAACTGGAAGAACGCTCGCAGAAGCGA AAAGAAGGAAAACAGGATGTAGAGTGC GCCAAAAGGGGGAGGAAAACGTGAAA ATTTTTCAGTTGCTAATTTTTCTTATAAA AAACAAAGTACTTTTAGGCATTCACCTG CATTATCTGAAACGTGGTTAAAAAAATA TCTTGTGCTATTGGCAAAACCTATGGTA ACTCTTTAGGTATTCCTTCGAACAAGAT GCAAGAATAGACAAAAATGACAGCCCT TCTACGAGTGATTAGCCTGGTCGTGA
	2.68	3.90	0.86	12.44	4E+06	STM3900					
	12.91	0.92	6.05	3.74	4E+06	STM3908					
	13.98	1.29	6.05	3.81	4E+06	IR STM3908 - STM3909	ilvY	positive regulator for ilvC (LysR family)	-		GGCCGAGATCTTCTTCCAGCCGCTGAA TCTGCCGGGAGAGCGTGGAGGGGCTG ACGTGCATCGCCCGCGCGCTGCGGCC AAAGTGGCGGCTTTCCGCCAGATGCAA GAAGGTTTTTAGATCGCGTAAATCCAC AGACAGACCTCCGGTTTTTGACGTTGC ATAAACCGCAACATAACGTTGTGAATAT ATCAATTTCCGCAATAAATTTCCTGTTG TAATGTGGGTTCATTCGCACAGATAGC AATCTGTAAACCGAACAATAAGCGCGA CACACAACATCACGGAGTACACCATCA TGGC
	18.44	2.07	7.27	1.04	4E+06	STM3909					
	4.88	2.98	3.83	2.83	4E+06	STM3945					

	2.89	3.25	2.76	2.32	4E+06	IR STM3945 - STM3946	STM 3945	pseudogene	-		AAAGATTGTTCTCCTCTTCTGGCTGGA GATAAACCACGCCGCTGCCTTGCCGCT GATAAACATTGTGCGGAGATTCACTCA GCCGGCATCCCCAGGCGGGAGGCAGC AGAAGTGAAAGCGAAAAAAGGCAAAAC AAATTACGATATTGCATAAGGTCATCCG GACGTGGTACGTAAACCTAAAGTGATG AGCAAAGCATGTTTCCTGATGTAAATG CGCAATAATCATGGCAACGCGCCGCTT TTCAGATTTTATAAAGAGCCCCTAAACG CTTGCTTTTACGCCTTCTCCTGCGATGA TA
	2.55	9.80	1.68	16.67	4E+06	STM3969					
	3.08	9.01	1.87	14.75	4E+06	IR STM3969 - STM3970	yigN	putative inner membrane protein	+		GGAACAGGCCGTTACGCAAGATGAAGA ATATCGTTTACGATCGATCCCTGAAGG GCGGCAGGATGAACATTATCCCAATGA TGAACGGGTGAAGCAGCAGTTAAGTTA ACCCATACGGAGTAGTTTAGTCCTGGC GCAGAGTAGGGCAAATTGGCCCAATCT GTTACACTTCTTGAACATTTTTATCGAT AAGCAGGCACTGAGATGGTGGAAGATT CACAAGAAACGACGCACTTTGGCTTTC AGACCGTCGCTAAAGAGCAGAAAGCTG ACATGGTGGCCCACGTTTTTCATTCTGT GG
	5.95	2.88	1.38	5.00	4E+06	STM3970					
	12.99	3.71	3.09	8.30	4E+06	STM4031					
	12.92	3.54	3.24	7.75	4E+06	IR STM4031 - STM4032	STM 4031	putative cytoplasmic protein	-		GTGAAGGAATATACCGCTTCATCTCTTC AGGCTGAGTGAATGTTTTTTTCTCCAGA ACATTCAGCAACTCAGTGAGAGCAAGC TCATGGTTTGGATACATGAGCATCGCT TCATTGAACGGTTTTCGGCTGATAACAT GCACAATGTAGTTCCATTACAAAGTTTT CAACCTGAAAACAATTTAGCGCAACGT TATCCAGTTTTCAAGTTGAAAACAAAAT TGAATTTTAGGTCATTTTGCCTGTTGAT GGACTTACAACACGCCAGGCCACATCT CGCATGGCGCTTCGTGCCGCCTGGC
	12.92	3.43	2.98	6.57	4E+06	STM4032					
	7.75	2.89	1.60	12.31	4E+06	STM4039					

	9.07	2.94	1.78	7.61	4E+06	IR STM4039 - STM4040	STM 4039	putative inner membrane lipoprotein	-		TACAGGTTGTTCGTCCGCTTTTTTTTCA TCACAAGCGCTTAGCCCGGCAGTCATC AGCATAGCGATAATAATTGATGATAACA AATCCTTTTTCATTAGAATAACCTATAAA TAATATCATTGAAATTTACAGATTCATTT TAATGAAAAAAAACAGGTATGTGATTTA TTCAACACAAAAAATACTTAATGCATAT TTCATTATAATTAACATTATCAATATCAA TGTGTTCGTTAAAATAAGAGAACCCCAA CGTAAATATACAAAAGGCAATTAAATGA AAAGGAATTTATTATCCTC
	7.72	4.08	5.59	16.26	4E+06	STM4073					
	8.23	5.98	5.62	12.28	4E+06	IR STM4073 - STM4074	yde W	putative transcriptio nal repressor	-		TCAATCCATCGTGATAGTAGAACCAGG CAATACGCGCCACCTGCTCTTCTTCGC ACATTCCATAATCAGATACCAACGTATT ATCGCTCATTGTCATAACCTGGCTTTAC TTTGAACATTTCTAAATCATTAACACAAT TGTTCAGTTATCACTCCGAAATAACCGT GATTAACGCCACAAAAACGCGCCAAAT CTGAACATTTATCATCTAAAAATTCATTT ATTCAGAAAACGTGATCTGGATGAGAG TTTTTTGACCAAATAACTACTACCGTTT TGAACAATTTCTTTTTCAAAAAA
	4.46	2.37	3.80	7.78	4E+06	STM4074					
	3.25	3.43	3.30	3.62	4E+06	IR STM4094 - STM4095	cytR	transcriptio nal repressor (GalR/LacI family)	-		CGCCTTCAACGCAACATCCTTCATCGT AGCGGCAGTAACCTGCTTGTTCGATTT CACTCTTTCTCCTCGCCTGGGAACTGC TGGCGCAGATCTATCCCTGGTAACACT CATCGAAAACATTTTTATCAGATAGTGC GTGGAAGCGGTTACAGAATTTTCATAA AAAGTGTGATGGATCTTTAATTTTACGA TCCGCCTCGCATCGTGAGGACTATCCT TCAATCGGATCGACGTCCAGAACCCAT TTAACTTTCCGCGCTTCCGGGAGCGTA TTGATCAACGCCAGCGTGCCGCTGATG AT
	5.79	3.45	4.28	5.46	4E+06	STM4095					

	11.08	5.52	4.05	11.01	4E+06	IR STM4111 - STM4112	ptsA	General PTS family, enzyme I	-		TGCCTTTGCGATCGGTGCGCAGGTTGT GCCACTCAATTTGCGACGTGAAGGTAT TACACAGCGTTTCTACGTGGCTTGCCG GGCGCGCATGTACGCCATTCGGCAGTT CACAGGTAAATTCCACAATCAGGGGCA TTGCCTCTCTCCCATAACGATTCTCTCG CTACAGCATAAAAGGAGGTAGCCGGAA TACGCCATGTGACAAATCTGTCAAAAG CTGGATAAATGTAATGTAGCGCAAAAA GTGCGAGTTGTCTCACAACTTAGCGTG GTAGCGCGGGTTTTACCTTTTTCAGAA GTT
	8.02	5.66	4.83	11.55	4E+06	IR STM4146 - STM4147	tufB	protein chain elongation factor EF-Tu (duplicate of tufA)	+		TTGGCGCGGGCGTTGTTGCTAAAGTTC TCGGCTAATCGCTGATAACATTTGACG CAATGCGCAATAAAAGGGCATCATTTG ATGCCCTTTTTGCACGCTTTCACACCA GAACCTGGCTCATCAGTGATTTTATTTG TCATAATCATTGCTGAGACAGGCTCTG TAGAGGGCGTATAATCCGAAAGGCGAA TAAGCGTTTCGATTTGGATTGCCTCGC GATTGCGGGGTGAAAATGTTTGTAGAA TACTTCTGACAGGTTGGTTTATGAGTG CGAATACCGAAGCTCAAGGGAGCGGG CGCG
	7.78	8.04	6.00	15.15	4E+06	STM4147					
	2.81	1.53	2.30	2.75	5E+06	STM4263					
	4.46	4.38	4.91	4.25	5E+06	IR STM4263 - STM4264	yjcB	putative inner membrane protein	-		TGTATTTTTTGTGCGTTTTATAACCGTA TTTTTTGTGTGACTTCTACGCGTCCGTA GAGAAACTGCCGGAAAGCAAAGATGTA TTATTACTACTCTTTTATTTTTTTTCGTG AAATTCAGACCTGATAAAAATATCAAGT TATTTATCAAAAGAAAGGAGTAAAGATG TATACCCCATCGTTTACTTGAGTATAAA TCTGATATTATCAAAAATATTTAGTGTC CTGCCTGGTATGCGAAAGAGATTGCGC GTAGTTATTAATGGTAAATGTTGATCGG TAAAAGTCTGTTGCTAATATTG
	2.64	9.15	5.09	10.54	5E+06	STM4326					

	2.72	9.21	5.11	11.48	5E+06	IR STM4326 - STM4327	aspA	aspartate ammonia-lyase (aspartase)	-		GCCACGCACAAATTCAGGGATGTCGCT GATTTTGTTATTGCTAATGTAGAAGTTT TCAATCGCTCTCAGAGTGTGAACACCA TAGTAGGCTTCAGCTGGAACTTCCCTG GTACCCAACAGATCTTCTTCGATACGA ATGTTGTTTGACATGTGAACCTTCTTTT TCAAGCTGCCAATGATTTTTACTTTAAA ACACACAGGATATATGTGATTTCGAATG TTTTCTGACCGACGATTATCCCCTCCAT CGGCCTGATAAACGAGATCATATGCTG GTTCAGAATTCCTACCGTAATCTGGA
	10.03	5.35	5.76	6.89	5E+06	STM4382					
	10.43	4.51	5.76	6.05	5E+06	IR STM4382 - STM4383	yjfR	putative Zn-dependent hydrolases of the beta-lactamase fold	-		GTACAGCCCAGCCACCACATAGCGAAC GTACCCGGCGCGACCTGCTCTTGTTCA ATCTCTTCGTTCAGCCAGCTTCCCCAC TCCGGAAACGTGCTCAGAATCCATGAT TCACGCGTGATGCTTTGTACTTTACTCA TCGCATTTACCTTCATGTTTGTTCAAAA TGGTTCAAAACGTGATTTGTTTTGATTA ATCCTGACACTATTTTCTCAAGAAGGCA ATGGGCTATTTTTTGACTTTTTGGAAGG AGAGAACGCAGTCAGGAGAAGATTTAA TCTTGTCTGGCGTCATGTGAATGTTT
	2.57	3.96	6.24	5.78	5E+06	STM4383					
	6.23	5.41	2.09	10.97	5E+06	IR STM4396 - STM4397	ytfB	putative cell envelope opacity-associated protein A	-		TTGGTTTTAATTCAAAGCGCCCGGGCA TGGTTTACCTCCTGCTCCGCATCTCGT TCCTTAATCATAGAGTATAGATGGCTAA CGCTATGATACTGGTAGTGCTATCCGC TTTCGTGACATCAATACGGATAATCTAT TGTTTCTTTTTCCCTGCGATTTGTCATC CTCCCTGAGACAAAGTTTTACCAGAAG AAGCGTGGCTGTTATGCTGCCCGCTAC TTTTTTGATATCCGATGAAGGAAAAATA ATGGCCACCCCGACTTTTGACACTATT GAAGCGCAAGCGAGCTACGGCATTGG T
	6.48	5.41	2.09	11.98	5E+06	STM4397					
	5.26	4.17	1.76	5.57	5E+06	STM4407					

	8.43	4.17	2.35	10.86	5E+06	IR STM4407 - STM4408	ytfL	putative hemolysin-related protein	-		TAATAACTTAAGTTTAATCTTACGTGAT GCGGCAAGCGAGATCTCGGAGATGGA GAAGAACGCACTTACAGCGATCAGGCA GAATATAATGAATATACTGTTTAACATA TCTTATCCGGCGAAACGCCAGATCCTC GGAAGGGAAGTTTATAAATCCGTGTGG TAACGTTTAATGAAAACCGGCTCGTAG CAGTGAGCCGATAAGTTCAGGGCTAGT ATAGCGTAAGCTACTGTAAAGTCGCCA GAGGGTTCATTTTCAACTCCGACAAGT TCCCCCTACGCCAGCGTCGTCACGCGT CAG
	7.16	3.68	2.35	16.47	5E+06	STM4408					
	16.03	2.44	1.33	7.29	5E+06	STM4408					
	23.39	2.09	0.54	6.79	5E+06	IR STM4408 - STM4409	msrA	peptide methionine sulfoxide reductase	-		CCCGAAAGCGTTAATTGGCGTTAAGGT TGTAACGAGACGCATCTTTGCACACAA TAACAACATTAATGTATCTGGATTTAAC CATAAGAAATATTTGGGCAGTCGTCTG CTTTTCAATCGAAATTGTTGATTTTATGT TAAGCCGCGGAGCGGTAGTGTGATTTT TTCCAGGGGTGGGAATAGGGGATATTC AGGAGAAAATGTGCCACATATCCGTCA GTTATGTTGGGTTAGCTTACTGTGCCT GAGCAGTTCTGCGGTAGCCGCAAATGT TCGTCTGAAAGTCGAAGGGCTATCCGG A
	23.39	2.11	0.59	6.79	5E+06	STM4409					
	9.38	2.77	1.77	6.46	5E+06	IR STM4416 - STM4417	mpl	UDP-N-acetylmuramate:L-alanyl-gamma-D-glutamyl-meso-diaminopimelate ligase	+		ACGTCATCTTCTGCCTTTCAACGTTTGC GATGCCGCCTGGCTGCGGGCATCGTC CAGTCATAACAATGCTGATCCTGTCGC ATTTATGCGGTCAGATTCAGATTGCTCA GAACCCAGCCCGCCAGCAAATTCTGTA CTGAAGGTAACCACAGCGCAATTTGAA TGTTGTTAACTGTATGTTCAGTTCATTT GTGCTAATATGGTTATTTACGAAATTTT CGTTCTATTAGAGTATCATGCATGTCTA AACATCAAACTCAACTTTCCTTACTGCA GGATGATATCCGCAGTCGCTATGACA
	9.63	3.11	1.87	5.93	5E+06	STM4417					
	3.07	3.12	0.52	4.64	5E+06	STM4473					

	3.19	2.34	0.42	4.90	5E+06	IR STM4473 - STM4474	yjgM	putative acetyltransferase	-		GGTAAGTCCGTATTCCGCTGAAACCTG ACGGATGACACGGGCAATAGCGGCATT GTCGGCGGTAGTGATTCGGCGCACCG TGAGCGTTGGCGAGGCGACATTATTCA TAATATGGCTCAATTTTTAAAATTTATTT ATAGATTACTTTAATACCACCGTCTTGA GTTACGCGCAAGGAGATCCTGAATCAG ACAAAATAAAAGGCGGAAAAATTAAACA AAAATAGTATCGTAGTCAAATCAGTAAC AGTTTACTGGTTTTTATTATTAATTCTAA TAGATTGTAATTCAGGGATATGATT
	4.42	2.41	5.25	6.54	5E+06	IR STM4501 - STM4502	STM 4501	putative cytoplasmic protein	-		TGTTCCTGACGGGATAAATTCATACTGA AGAACCTGTTTAATCATCATAGGCTAAA CGTGCAAACACACTGCGGTGTCCGCAT TCGATTTCGGCGCATTGATAATCAGTC CGGCCTGAAAAGGTCGGGTAACTGATT ATCAGATGATGACATTCTCCAGCATCAA AGCCTCGGGTTGAGTTGAAAGGTATTT ACGTCGTGAATGATAACACCTGATTTCT GTAAGTGAATAACCGGGAGTGAAAAGT GTGATCTCAAAGGGAGGCTCATGACGT TTAGCGTATCAGATGAATAGCTCCCGC

Table 3B Regions that induce GFP expression in both tumor and spleen (cont'd, presented in the same order as Table 3A)

3' gene	Function	3' gene orientation
STM0649	putative hydrolase N-terminus	+
hutU	pseudogene; frameshift relative to Pseudomonas putida urocanate hydratase (HUTU) (SW:P25080)	+
STM1056	Gifsy-2 prophage; homologue of msgA	-
STM1265	putative response regulators consisting of a CheY-like receiver domain and a HTH DNA-binding domain	+
ydgF	putative membrane transporter of cations and cationic drugs	+
pspD	phage shock protein	-

STM1698	putative inner membrane protein	-
nhaB	NhaB family of transport protein, Na+/H+ antiporter, regulator of intracellular pH	+
STM1839	putative periplasmic or exported protein	-
yegE	putative PAS/PAC domain; Diguanylate cyclase/phosphodiesterase domain 1, Diguanylate cyclase/phosphodiesterase domain 2,	+
cdd	cytidine/deoxycytidine deaminase	+
yfgB	putative Fe-S-cluster redox enzyme	-
gshA	gamma-glutamate-cysteine ligase	-
deaD	cysteine sulfinate desulfinase	-
hopD	leader peptidase HopD	+
pckA	phosphoenolpyruvate carboxykinase	+
ftsX	putative integral membrane cell division protein	-
yhjS	putative cytoplasmic protein	+
STM3624A	putative protein	+
rpmH	50S ribosomal subunit protein L34	+
cyaA	adenylate cyclase	+
udp	uridine phosphorylase	+
yiiU	putative cytoplasmic protein	+
rsd	regulator of sigma D, has binding activity to the major sigma subunit of RNAP	-

ecnB	putative entericidin B precursor	+
ytfF	putative cationic amino acid transporter	-
ytfK	putative cytoplasmic protein	+
idnK	D-gluconate kinase, thermosensitive	+
STM4552	putative inner membrane protein	+
deoC	2-deoxyribose-5-phosphate aldolase	+
PSLT048	alpha-helical coiled coil protein	+
djlA	DnaJ like chaperone protein	+
stfA	putative fimbrial subunit	+
frr	ribosome releasing factor	+
uppS	undecaprenyl pyrophosphate synthetase (di-trans,poly-cis-decaprenylcistransferase)	+
yaeQ	putative cytoplasmic protein	+
STM0307	homology to Shigella VirG protein	-
STM0341	putative inner membrane protein	+
STM0343	putative Diguanylate cyclase/phosphodiesterase domain 1	+
phoB	response regulator in two-component regulatory system with PhoR (or CreC), regulates pho regulon (OmpR family)	+
cypD	peptidyl prolyl isomerase	+
ybaY	glycoprotein/polysaccharide metabolism	+

acrR	acrAB operon repressor (TetR/AcrR family)	+
aefA	putative small-conductance mechanosensitive channel	+
cysS	cysteine tRNA synthetase	+
fepE	ferric enterobactin (enterochelin) transporter	+
cobC	alpha ribazole-5'-P phosphatase in cobalamin synthesis	-
kdpE	response regulator in two-component regulatory system with KdpD, regulates kdp operon encoding a high-affinity K translocating ATPase (OmpR family)	-
STM0763.s	transcriptional regulator	-
STM0835	putative Mn-dependent transcriptional regulator.	+
STM0860	putative inner membrane protein	-
yljA	putative cytoplasmic protein	+
STM0947	putative integrase protein	-
lrp	regulator for lrp regulon and high-affinity branched-chain amino acid transport system; mediator of of leucine response (AsnC family)	+
serS	serine tRNA synthetase ; also charges selenocystein tRNA with serine	+
ycaO	putative cytoplasmic protein	-
STM1001	putative leucine response regulator	-
STM1020	Gifsy-2 prophage	+
sulA	suppressor of lon; inhibitor of cell division and FtsZ ring formation upon DNA damage/inhibition, HslVU and Lon involved in its turnover	-
copS	Copper resistance; histidine kinase	-

ycdF	pseudogene; in-frame stops following codons 5 and 21	+
rluC	23S rRNA pseudouridylate synthase	+
potB	ABC superfamily (membrane), spermidine/putrescine transporter	-
STM1263	putative periplasmic protein	+
yeaR	putative cytoplasmic protein	+
celA	PTS family, sugar specific enzyme IIB for cellobiose, arbutin, and salicin	+
ydiM	putative MFS family transport protein	-
ydiJ	paral putative oxidase	+
pykF	pyruvate kinase I (formerly F), fructose stimulated	-
orf242	putative regulatory proteins, merR family	-
ydhL	putative oxidoreductase	+
malY	pseudogene; in-frame stop following codon 16	-
ydgC	putative inner membrane protein	+
yncC	putative regulatory protein, gntR family	-
ynaF	putative universal stress protein	+
adhE	iron-dependent alcohol dehydrogenase of the multifunctional alcohol dehydrogenase AdhE	+
hnr	Response regulator in protein turnover: mouse virulence	-
STM1786	hydrogenase-1 small subunit	+
STM1795	putative homologue of glutamic dehyrogenase	+

minC	cell division inhibitor; activated MinC inhibits FtsZ ring formation	+
yobG	putative inner membrane protein	-
STM1841	putative outer membrane or exported	+
STM1856	putative cytoplasmic protein	+
pagK	PhoPQ-activated gene	+
STM1934	putative outer membrane lipoprotein	+
fliB	N-methylation of lysine residues in flagellin	-
STM1967	putative 50S ribosomal protein	+
STM2148	putative periplasmic protein	+
yehV	putative transcriptional repressor (MerR family)	+
yohJ	putative effector of murein hydrolase LrgA	+
yejL	putative cytoplasmic protein	+
STM2281	putative transcriptional regulator, LysR family	+
yfbQ	putative aminotransferase (ortho), paral putative regulator	+
yfcX	paral putative dehydrogenase	-
nupC	NUP family, nucleoside transport	+
yffB	putative glutaredoxin family	+
ndk	nucleoside diphosphate kinase	-
hmpA	dihydropteridine reductase 2 and nitric oxide dioxygenase activity	+

gogB	Gifsy-1 prophage: leucine-rich repeat protein	+
STM2621	Gifsy-1 prophage	-
nadB	quinolinate synthetase, B protein	+
yfiO	putative lipoprotein	+
ygaM	putative inner membrane protein	+
proV	ABC superfamily (atp_bind), glycine/betaine/proline transport protein	+
hilD	regulatory helix-turn-helix proteins, araC family	+
STM2904	putative ABC-type transport system	+
STM2954.1n	hypothetical protein	-
kduD	2-deoxy-D-gluconate 3-dehydrogenase	-
yohM	putative inner membrane protein	+
ygfE	putative cytoplasmic protein	+
rpiA	ribosephosphate isomerase, constitutive	-
STM3084	putative regulatory protein, gntR family	-
STM3169	putative dicarboxylate-binding periplasmic protein	+
yqiC	putative cytoplasmic protein	+
ygiM	putative SH3 domain protein	+
yqjI	putative transcriptional regulator	+

rnpB	regulatory RNA	+
yhbY	putative RNA-binding protein containing KH domain	+
STM3343	putative cytoplasmic protein	-
STM3357	putative regulatory protein, gntR family	-
accB	acetylCoA carboxylase, BCCP subunit, carrier of biotin	+
def	peptide deformylase	+
slyX	putative cytoplasmic protein	+
hofQ	putative transport protein, possibly in biosynthesis of type IV pilin	-
yrfF	putative inner membrane protein	+
feoA	ferrous iron transport protein A	+
gntT	GntP family, high-affinity gluconate permease in GNT I system	+
livF	ABC superfamily (atp_bind), branched-chain amino acid transporter, high-affinity	-
uspA	universal stress protein A	+
STM3631	putative xanthine permease	-
mtlA	PTS family, mannitol-specific enzyme IIABC components	+
STM3794	putative regulatory protein, deoR family	+
torD	cytoplasmic chaperone which interacts with TorA	-
STM3858	putative phosphotransferase system fructose-specific component IIB	-
ilvL	ilvGEDA operon leader peptide	+

ilvC	ketol-acid reductoisomerase	+
yifL	putative outer membrane lipoprotein	+
ubiE	S-adenosylmethionine : 2-DMK methyltransferase and 2-octaprenyl-6-methoxy-1,4-benzoquinone methylase	+
STM4032	putative acetyl esterase	-
yiiG	putative cytoplasmic protein	+
ego	putative ABC-type sugar, aldose transport system, ATPase component	+
priA	primosomal protein N' (= factor Y) directs replication fork assembly at D-loops	-
frwC	PTS system fructose-like IIC component	+
secE	preprotein translocase IISP family, membrane subunit	+
yjcC	putative diguanylate cyclase/phosphodiesterase	+
fxsA	suppresses F exclusion of bacteriophage T7	+
sgaT	putative PTS enzyme IIsga subunit	+
fklB	FKBP-type 22KD peptidyl-prolyl cis-trans isomerase (rotamase)	+
msrA	peptide methionine sulfoxide reductase	-
ytfM	putative outer membrane protein	+
STM4417	putative transcriptional regulator	+
yjgN	putative inner membrane protein	+
STM4502	putative cytoplasmic protein	+

Table 4. Intergenic regions that induce higher GFP expression in spleen than in tumor

Clone ID	Spleen	Tumor (+)	Tumor (+)(-)(+)	Tumor (+)(-)(+)	Genome position of peak signal	Gene	Gene symbol	Gene orient.	
	lib1	lib2	lib3	lib4					Sequence
	Median of experiment versus input library								
	lib-1	lib-2	lib-3	lib-4					
	moving median of 10	moving median of 10	moving median of 10	moving median of 10					
	16.24	0.84	0.41	0.37	7389	STM0006	yaaJ	-	
	22.42	1.98	0.38	0.33	7513	IR STM0006 - STM0007			GTATTTCGTTAATAAAACTGAAAAAC TCAGGCATTAACGTCCCTCTTGTTG ATGCCGGCACGCTTTGATAATCCTG TATAAGCGTGACCCATGATGTAGAT GACCTTGTCAGACTAATATTAACGG CAGTTTACCATAAATACGGTGGTAT CCTTTAATTGCGCATCAACCGTCGG CAGATACGCAAACAGTGCACAAGG GCAGCCAGGTGCATGTAGGCGGTT GCGCTGTGAGTGCGTCGTGTTATCA TCAGGGTAGACCGGTTACATCCCCT AACAAGCTGTTTAAAGAGAAACTCT AT
	21.01	1.73	0.38	0.30	7662	STM0007	talB	+	
	1.58	0.92	1.20	0.38	93836	STM0080		+	

	20.94	0.46	0.93	0.29	94051	IR STM0080 - STM0081			TGCGAATAAACGGATGCCTGAACAG GCAGGGACGCCGGAAAACGTCGAA ATACGTTAGACCATTCGCCCGTGTT CCCGCTTTCCCCACCGCGCTGTCC GCTTACATGAGGTTACACTCATCGA CATTTCTCTGAACAGCGGCTCAACA TTTCCCGGAAAAAAACATATCGCAG GGCATTTATCCTTATGATTAGGTATA AATGATGAGGTATAAGGAACAGGAG TCTGTAATGAAACCAATACCTTTTTA TTTGCTCGCGCTATTTTCTGCCGCC TCCGGGGCTACGGAGATAAACGTC TG
	25.94	0.56	1.06	0.31	94098	STM0081		+	
	17.77	1.63	2.35	0.31	442273	STM0390	aroM	+	
	14.65	0.81	0.65	0.28	442548	IR STM0390 - STM0391			TCAAGGCGCGGACGTCATTATGCT GGATTGTCTGGGTTTTCATCAGCGT CATCGGGATATTTTACAGCAGGCGC TGGATGTGCCGGTTTTACTCTCTAA CGTTTTGATTGCGCGGTTAGCTTCA GAACTGCTTGTCTAATTTTACGTGA CAGGCCGAACGTCAGGACTCTATAT TGGGTGTTAATTTAATAATGAGACG GGGCCTGATTATGCTACAAAGCAAT GAATACTTTTCCGGGAAAGTTAAGT CTATTGGATTTACCAGCAGTAGCAC CGGCCGGGCCAGCGTTGGTGTGAT GGC
	8.00	0.73	0.68	0.29	442570	STM0391	yaiE	+	
	9.82	1.66	0.42	0.52	667851	STM0605	ybdN	-	

	9.82	1.76	0.43	0.61	667878	IR STM0605 - STM0606			CAACGTTGCCGTCAGGTGCAACATA AGTCCTGAATCTTTACCACCAGAAA ATGAGACGCAGACCCGGGGTAAGG TTTCCAGGGTCCACATTATACGCTC TTGAGCCGCTTCCAGAACATTTTGC TCGAGCGGAACTTTATAAACCGACA TCTCTGGATAGTCTCCGATGTGTTA ACTACAGTATATTCGAAATAATTAAC ATAAAGGATAAGCAGATTAGATGAA CTTGCAATGCTTTATTATATTTGTAA AATAAATATATTCCATAAACATATAC ATTAAATTTATATTAATATCCGTT
	4.72	0.66	0.90	0.70	668757	STM0606	ybdO	-	
	15.90	0.66	0.71	0.25	962476	STM0892	ybjP	-	
	10.80	0.44	0.63	0.31	962530	IR STM0892 - STM0893			TGAGCCACGCTGTCCGGGCCGCCT TCCACACACGCGCCGATACGCGGG CCATTATCTTTGTAGGCGGGAGTGA CGGTCGTACAGGCGCTAAGCAGAA GCGCGCACGGGATGAGCAAAGAGA GTTTAGAATAGCGCATGATGATTTC CTTATAGGCGATCGAGCAAAAACCG ATCTACGATAATCAATTATATCCTTT CAGTGATTGCATAACCACTTAACAT CTTGTTTTATCTAAATAAAATTAAGC ATGTTATCTTTTTGGGGCACTCCTG GGGCAGTAGATGCCAGTTGTTGATT CAG
	6.64	0.41	0.75	0.58	962570	STM0893		-	
	5.69	0.32	0.27	0.39	1E+06	STM1044	sodC	-	

	8.09	0.63	0.32	0.39	1E+06	IR STM1044 - STM1045			ATGTTTTCTCCTGTTCCGCTGGACA GGGCATCGTTCATCTTTACAGTCAG GGTATTCTCTGCCATTGCTGAACAA CTGATGAGCGCACCAGCTACCAGC GACAATATTGTGTATTTCATTAGTTA CCTCGTTTTTTGGTTGTATCGTAAAT ACCATTAATAAAAGCAGGTATATGTT TGCAAGATAAATAATAAAGGATCTC TCATATATGCAGGATATACCACAGG AAACCCTGAGCGAGACCACCAAAG CGGAGCAGTCCGCGAAGGTGGATT TGTGGGAATTTGATTTAACCGCGAT T
	10.05	0.88	0.38	0.50	1E+06	STM1045		+	
	12.79	0.74	1.01	0.23	1E+06	STM1231	phoP	-	
	12.76	0.74	0.45	0.23	1E+06	IR STM1231 - STM1232			AGGTGTTCATTAAGGTAGTAATCAG CTTCCCTGGCATCTTCTGCGGCATC GACCTGGTGACCTGAATCCTGGAG CTGAACCTTCAGGTGGTGGCGTAAT AATGCATTATCCTCTACAACCAGTA CGCGCATCATCTCTTCTCCCTTGTG TTAACAATAAGAACAGTCTAGCGTT GATTATGGTGCTTTGGGGATAAACA GTTAATAAACCAGACAAATAGTCAC CCTCTTTCTGAAGAAAAGAGGGTGA GGCAGGCATTATTTAAGTTCGTCGA CCAGAGTCACAGCGCGACCGATAT AAT
	9.96	0.61	0.45	0.30	1E+06	STM1232	purB	-	
	1.16	2.63	6.81	5.31	1E+06	STM1249		-	

	31.95	0.64	1.01	0.40	1E+06	IR STM1249 - STM1250			TCAGTGAAACTATTTCTTCAAATGAT GGTCTTTTTATTATCGATCAGATAAT GGCATCAACAGGGGTTATTCAGGA GTATATGTGAAAAAGTGGCTTATAG GAGGGATATTGATCGCAAGTTTTCT GACCGGTTGTCTGATGTGGCACAA CATTGATAAATGGTTTAATAAAGATA TCGAATTTTTCTACGTCGGAGACGA TAGCTAAAATTCCAGTCAGTTGGCA ACGGGTGTCATATCTTCAGGTATGG CGCCCGGAGCCGCCGGGCGCAAAT TGTAGGTGTATAAAAGTCATTTCATT
	12.37	0.82	0.82	0.48	1E+06	STM1250		+	
	11.46	1.34	0.41	0.33	2E+06	STM1583		-	
	10.52	1.60	0.34	0.44	2E+06	IR STM1583 - STM1584			TGCGGTAAGCACATACAAGATGCCT TTCATGATTTTTGTTGATAATTTATTT TCATAATCTCCTGCAGCAACATGAG GTAGCTTATTTCCTGATAAAGCTCT GGCATAGGTAGAAACTGATGTATAT GGCATATCCTACTCCTTCAAATTTTG CTCAATAGCTTTATATGTCCTACTCC TCTCTCATTATGACGATATGTCAATC AACAAAATTGCTCAAAGGCATACAT TTTCAGGAGAAAATGAGAATAACAG GCGCAACGGCCTGATCTTATGCTG CTTCAATATCGTCAGGTGGTTT
	2.44	0.56	0.92	0.41	2E+06	STM1584	ansP	+	
	34.34	1.01	0.56	0.26	2E+06	STM1736	yciA	+	

	38.32	1.01	0.57	0.29	2E+06	IR STM1736 - STM1737			ACGACGTCTATTAGCATAAATATTG AAGTCTGGGTGAAAAAAGTCGCGTC AGAACCGATTGGGCAGCGCTACAA GGCCACCGAGGCGCTGTTTATTTAT GTTGCCGTCGATCCGGACGGTAAA CCTCGCCCGCTCCCGGTTCAGGGT TAAGTATACCCGCTTACGCCGCCAG CAGGTGATGGTATATTCCTGGCTGG CGGCGCCAGAGATTACTCAATCTGC GCCGTACCGTTCAGACGGAAGATA ATATTGACCACCAGCCCGGAACCC GGCTTGCCTGCTTCATAGCGCCATT TTCGCA
	39.25	0.95	0.69	0.30	2E+06	STM1737	tonB	-	
	1.31	1.19	2.93	0.37	2E+06	STM1868.1N		-	
	10.59	1.46	0.38	0.48	2E+06	IR STM1868.1N - STM1868A			GTTCGCCGTCCATTTTTACCTCTGG GGCTGTTTCTTAGCGCGCCCTCCC CCGGAAAAACAAAATATAATGAACA AAAAACATACAAACCATCATCTTTTA AAAATAAATTACATTAAAACAGAGAG TTACAACATGATGATGATGCATGAA AAATCAAAAATGCGCCAAATCCCGC GCCGCTGCCGCCCCGTGGCAGGC CGCCCCGCCGGGAGTACCTTTTTAA AATGCGAACAATTATCAACAACTAC CACTTAATGATTATTTATTTCATTTT GCGATATTGATTATCATTTTCAATAA
	8.17	1.52	0.22	0.31	2E+06	STM1868A		+	
	11.80	1.45	0.68	0.33	2E+06	STM1876	holE	+	

	14.81	1.25	0.83	0.34	2E+06	IR STM1876 - STM1877			GCTACAATATGCCAGTTGTCGCGGA GGCGGTCGAACGTGAGCAGCCAGA GCATCTACGCGCCTGGTTTCGCGA GCGGCTGATTGCCCATCGTCTGGC TTCCGTATCACTATCCCGACTCCCT TACGAACCCAAAGTTAAATAAAAATT ATATAACGTTACACTTCCTTACATGC AGACGACTACATTATAAGGCGATTC TTAACCTATGCTTTTTAGAATGGCTG TAGAGACTATGAAAAGGAAGTCATT ATGTCCTCCTGGAAAATTGCTGCTG CGCAGTATGCGCCCCTGAACGCCT CG
	12.07	0.81	0.97	0.37	2E+06	STM1877		+	
	14.41	0.62	0.43	0.33	2E+06	STM2153	yehE	-	
	19.07	0.61	0.39	0.37	2E+06	IR STM2153 - STM2154			GGTTAATGTTGCGGTGTCGGAGGC AAAAACAGGTACGCTTATCCCATAA GCCGAAACTATAATTCCCATCAGCA AATATTTTTTCATAGTGAGTAATTGT TCCTCTGGTGAACGTCAAACAGTAT GCAGGCCGTCCTGATGAGCAGTAT GAACGTATCGATACCTTAAAACCAA TTGAAAAAATAAATCAGTAGGATAG GTATGATCAATTCAAATAATGTTTTT GCCGATTATTTCAGATAAACACCTG TCTGTTTAAGCAGGAATTAACAATG CGGGGGCTATTATTTTATTAATACAT
	4.64	1.02	0.57	0.41	2E+06	STM2154	mrp	-	
	11.33	1.37	0.82	0.45	2E+06	STM2169	yohC	-	

	11.99	1.53	0.81	0.45	2E+06	IR STM2169 - STM2170			ACGACGGGAATCGCCGCCATCAGC AAAACATGGTGCGTATAGTGATGCG AAACAGTTTCGTTTTCGCTTTTGATC ACCTGCATTTCCCGATCGGGATGG GAAAAAAGCCCCCATACATGGTTCA TACTGCCCCCTTCTGCTGCCTCAGA TGCCAGTATGTTCAAGTATAATTCA GTTTCTGGTTATTTTATGAACAATGG CAAAATAGTCTCCGGCAAAACGTCG GCTTTGCCGCGCACGCCTCTTGCC AGGGTGTATGCTTAATGCCGGAGG TGGTTTACGCATGGATATCAACACG CTT
	11.13	1.58	0.80	0.47	2E+06	STM2170	yohD	+	
	20.97	0.90	1.83	0.42	2E+06	STM2349	yfcG	+	
	17.50	0.66	1.54	0.33	2E+06	IR STM2349 - STM2350			GATCTTGATACCTACCCGGCGGTGT ATAACTGGTTTGAACGCATTCGCAC GCGTCCTGCGACAGCGCGCGCACT GTTACAAGCGCAACTGCACTGTAAC AGTACGAAAGCGTAACGCGGTAGC ATACATCATGTATGATGTAGAGGTG TATACACGGAAAAAACCTGCGTCCG GCACCCTTATTCGTATTAAAAACCT GACATTAGGGAAGAGGAAATCCTCC CTACTCTGGAGGTCATATGCAGATT CTGATTACCGGCGGTACAGGCCTG ATAGGGCGTCATCTCATTCCCCGGC TGTT
	13.83	0.67	1.52	0.33	2E+06	STM2350	yfcH	+	
	14.01	1.14	1.19	0.43	2E+06	STM2366	accD	-	

	11.78	1.29	1.15	0.39	2E+06	IR STM2366 - STM2367			CTCAAGATTACGTTCCAGCTCAGCG CGGTATAAAACCTGACCGCAGCTAT CACACTTGGTCCACACCCCTTCAGG AATGCTAGCCTTGCGGGTGGGAGT AATGTTGCTTTTAATTCGTTCAATCC AGCTCATTGGTGACCTTTCTGCCTG AACCTTAGTCAGCTTTATTATAAGG GGCGCATAATGCCATTTTTGCCCCC AACAGACCATGAATGTTGCACATTA AAACATAACAGCCCGAAACTTTGGA TAAAAAAGTGGTCGAACCGCTGAGT TACTTTCTATTTTGCGGCACGCGAC G
	3.49	0.92	0.89	0.35	2E+06	STM2367	dedA	-	
	1.89	0.55	0.31	0.26	3E+06	STM3047	ygfY	-	
	10.99	0.73	0.24	0.26	3E+06	IR STM3047 - STM3048			ATTGTGAATATCCATGTTCTTCCTGC CTCGCGAAAATGAAGTACCGGGCT ATTGTAACGTGTTTTTGGCGTTGTTT TACGGGAATCTCAGTAATCTGGAAC GCGATCGCGAAATAAAAGGCTGGG AATCAATATGTTCATCCATTTTGGAT ACCGCCTCGCAAAACGATCAATCCG CTCTCAATGGGCTATTTAAAGCACT TGCAATGACCGATGGCTCTTTTACC ATTAACCATTATTGTTGCAGCTAACC AGGACATTATTTATGGCTTTTATCTC CTTTCCACCACGTCATCCTTCAT
	12.16	1.18	0.31	0.30	3E+06	STM3048	ygfZ	+	
	9.40	0.58	0.91	0.42	3E+06	STM3231	yqjK	+	

	14.81	0.63	1.13	0.54	3E+06	IR STM3231 - STM3232			GGTCGGTAGCAGCGTAATGGCCAT CTGGACCATCCGTCATCCTAATATG TTGGTACGCTGGGCGAAACGCGGC CTGGGTATCTGGAGCGCCTGGCGC CTGGTAAAAACTACCCTCCGTCAAC AACAGCTCCGCGGTTAATATCTTTT CTTTTATAGCATCGCGCCATCAGGT TATCACCTGGTGGCGCGATACTTTT ATGCATATCGTCTCTTTAGCAATCA CTCAAATTTTTTGAAAAAATTTGGCA ATTTTCCTTGCTAACAATTCCTGCAC GCCACGTTTATGATTCTCTCCAGCG AT
	11.41	1.09	1.30	0.41	3E+06	STM3232	yqjF	+	
	2.83	0.88	1.96	0.25	4E+06	STM3805	yidH	-	
	10.53	0.55	1.90	0.28	4E+06	IR STM3805 - STM3806			GACGCCTGCCGCCAGAAATCCCAG CGAGGTGCGAATCCACGCCAGAAA GGTGCGCTCATTTGCCAGTGAGAA GCGATAATCCGGCGCTTCTCCGAG GCGGGAAATCTTCATGACGACTCCT TTTACGTTCTTATGTATTCCCGTTCG TTTTCAGAATACCACTCACGTTGTT GCTGATATGCTTCACATTATCCCGC AGCAAGGGAATCTTATTGCAAAATA ACTGTAGTTCACTGGTGATGCGTTT TGGCGCAACCGCGCTCATTGCCGC TATTTTTCATTTCAGTTACGACCTTT TTCA
	14.49	0.95	0.95	0.37	4E+06	STM3806		+	
	3.74	1.05	0.59	0.26	5E+06	STM4286	lpxO	-	

	9.12	1.26	0.50	0.36	5E+06	IR STM4286 - STM4287.S			CGGTGATGCCAAAGAGAAAAGTGTA GTTCGTTGACAATAAATTTACATTTC TACAACTTAAAAGGGCCATTTTTGC TAAAGAAGCGAGTCAGCCCGTTTAA CCTTTATCCAGGCTTGTCGACAGTA GAATTGAGATGACTCCGCTACTTCA CCCGGTGATGGCTGATTACGTTATG CCTTATCTCCCGATGACGGCTGCCA GATCACAATGCTTTCGTAAACCGAA AATGACTTTGCTTGTAACCTTCGCG AAGATAAAAACGGTGTGCATCGCG GCGTTTAATATTTGTGGAAAGCTCC G
	9.12	1.29	0.50	0.36	5E+06	STM4287 + STM4287.S			
	7.62	1.72	0.64	0.41	5E+06	STM4290	proP	+	
	7.69	1.57	0.62	0.41	5E+06	IR STM4290 - STM4291			GCGTCGGACATCCAGGAAGCGAAG GAAATTCTGGGCGAGCATTACGATA ATATTGAGCAGAAAATCGACGACAT CGATCAGGAAATTGCGGAGCTGCA GGTCAAACGTTCGCGTCTGGTACA GCAACATCCGCGTATCGATGAATAA ATTTCGCGCTTAAGGTTCGCTTAAT CTCTCGCGGGCATACTCTCCTCCAT ACCTTTGGAGGAGAGCGTCATGAAA AGCTATATTTATAAAAGTTTGACGAC CCTGTGTAGTGTGCTGATTGTCAGC AGTTTTATCTATGTGTGGGTCACGA CGT
	1.41	0.75	1.79	0.35	5E+06	STM4291	basS	-	
	18.03	1.30	0.20	0.27	5E+06	STM4328	yjeH	-	

	17.61	1.11	0.22	0.30	5E+06	IR STM4328 - STM4329			GATGTGGTTAACAAGATAACGCCCT GAACCAACCCAAGCTCTTTTTTTAG TTCATTCATCAGCTCATTATCCGGC GGCATTGTAACGTCAGGTGACGAC AGACATTTTTAAGCGTATCACACAC GCCTTTTCTTATAGCAGGATGTTCT AAACCTTGGGTAAACGTGAGATAAG TAGCGTTTTTACCGCTTTTTTCGCTC AGAAGAATTTTTTTTTCATCTCCCCCC TTGAAGGGGCAAAACCCCATCCCC ATCTCTCTGGTCACCAGCCGGGAAA CCGTTTACGGGCCGGCGTCACCCA TA
	2.21	1.06	0.57	0.48	5E+06	STM4329	mopB	+	
	28.58	0.84	1.28	0.56	5E+06	STM4362	hflX	+	
	35.05	1.86	1.16	0.37	5E+06	IR STM4362 - STM4363			AGCGTCAGTCTGCAGGTACGAATG CCGATTGTCGACTGGCGTCGCCTC TGTAAACAAGAACCGGCGTTGATCG AATACGTGATCTAGACGCGAAGTCA TTCAGGTCGTATTGAGGCGGTAGCT GGAGAGAATCTCAGGAGCTCACAA CGAAGTGACCTGGGGTAAAAAAGC CGCCACTCAAGACGCAGCCTGAAA GATGATGTCTGTAACGGCGGTTCGT CTGAAGCATGGAGTAATTTCGCCTT ATCCTCTGAGGTCGAAAGACAACG GGGATCACCGCATAACAAATATGGA GCACAAA
	33.31	0.91	1.01	0.29	5E+06	STM4363	hflK	+	

	9.82	0.90	1.26	0.48	3113	IR PSLT006 - PSLT007			AAACTGCCGCCGGAGCCGCGTGAA AATATTGTTTATCAGTGCTGGGAAC GTTTTTGCCAGGCATTGGGGAAAAC CATCCCGGTGGCGATGACGCTGGA AAAAAATATGCCGATTGGTTCCGGG TTAGGGTCCAGCGCCTGTTCCGTC GTCGCCGCGCTGGTCGCGATGAAT GAGCACTGCGGCAAACCGTTAAAC GACACGCGTCTGTTGGCGCTGATG GGCGAGCTGGAAGGCCGTATCTCC GGCAGCATCCATTACGATAACGTCG CGCCGTGCTTTCTTGGCGGTATGCA GTTGATGA
	2.88	0.48	0.74	0.34	3721	PSLT007		+	
	7.69	0.92	1.67	0.45	17888	IR PSLT024 - PSLT025			TCATTTTTATGATTTTTATATCATCTA AAAAGATGATGTTTTGTGATTAGCTA TTTTTTATGCCTGTAACGATTATGGA CCCCGCAGAACGAGCTGCGACAAT TTTGAAACGTAAAAGGAAATTTGAA AATGGCTACAAGCAAACTGATTCAA GGCGATACAATTACTGAAACTACTC ATGCAGCGAATGGTTTTGACCCTGC AACAAGCGATGATAAAATAAGCTAT ACTTCCGCTCGTGTTGCGAAACCG GTATACAATAAATATAAAAATTCCAC GACTAAACCGAAGGTATTCGGTT
	5.19	0.66	1.53	0.40	18097	PSLT025		-	

	3.20	1.01	0.82	0.38	18666	IR PSLT025 - PSLT026			AACTGTTCAAACAGTTCCCGATGTT CAGCGAAGTGGATATTGACTGGGA ATACCCGAACAATGAAGGGGCGGG CAACCCGTTTGGTCCGGAAGATGG CGCTAACTACGCGCTGCTGATTGCC GAACTGCGTAAACAGCTGGATTCCG CGGGTCTGAGCAATGTGAAGATCTC TATTGCCGCTTCTGCTGTCACTACT ATTTTTGACTATGCGAAAGTAAAAG ATCTGATGGCTGCCGGCCTGTATG GCATCAACCTGATGACCTATGACTT TTTCGGTACGCCGTGGGCGGAAAC GCTGGG
	3.84	1.29	0.49	0.36	30863	PSLT040	spvA	-	
	12.30	0.93	1.84	0.37	31227	IR PSLT040 - PSLT041			CGTGGCTCCCTTTGCAACGCGTCAA ACGGACTGGTGCCGGCACACGGTT CGCTGCACTGTGCGCTGGCAAAGT ATTAATGACTATGGGCGGGTAATGC CAGCGCAAACCGTGGATCTGACGC GTATTCATTAACCTATTTTTCAGGCG TCTCCCGATAGCGGGAGGCTTTCC GAACTTATCGAACGAGACTTTTATTA TGTATTATCACGCGTTAAAACTTTCC CGACTGGCGATGTTGACGTTGGCA GGCGTTGCCGTATCCGCCTCGGCA ATCGCCGCCGATTCTGCCCCGACG TCGCA
	7.27	1.02	3.20	0.51	31383	PSLT041	spvR	-	

	7.16	0.55	1.08	0.74	32347	IR PSLT041 - PSLT042			TCCTTTATCGTTCATGAAGGGACAG CGAAACCGACCGCTCAGATTCATTT TATGGGATCGGTTGTTGAGGCAGG CTGCTGGAATGACGTAGGAACCTTA GAAATTCAATGCCATAATAAAGAGG GAGTTGAACGTTATATTATTGTCGA GAATATTATCACGCCGATATCGTCT CCTCATGCAACGGTAAAACGAGATT ATTTGGATGAAGATAAGCAATTAAC AGTGCTACGCATTGTCTATGACTGA ACCGCGTAGCAGACCGCAGATGGT GTCCCGTCAGTGTCGTGTGAGAATA TTA
	11.80	1.53	1.25	0.51	35187	PSLT044		-	
	2.87	1.13	1.28	0.40	37474	IR PSLT045 - PSLT046			CAATACGCTGGCCCAGCGGTTTGG TGCTGTCATATTTAAACTGGACGGT TTTAGATACGTGCAGCATACCGTTT TTCAGATCGGCAGCGTGTGACATGA TGGATTTCAGGTCCTTACCGCTGAT TTCCATGCTCATGACATCGTTGGTG AACGGATACATACTCAGCACATCAC CATAGGTGATATTACCTTTAGGCAA TTCGGTACGGATGCCGCCAGCATTA TAGAAGGAAGCGTCGGCGCCAGGA ACGGTAGCCATCAGGGCATCGGTG ATTAAGTTGCCGGTTGGCGCGGATT CACC
	10.57	1.16	0.91	0.60	38107	PSLT046		-	

	5.16	1.15	1.60	1.64	38398	IR PSLT046 - PSLT047			CATTATCCAACAATACCGGGAATTG CAATTTGCTGAGTTGTTTAACCAGA TTCTCATGGCCATGGTCAAATTCAT GGTTACCGACAGAGACGGCGTCGT AAGGCATGGTATTTAAAATATCAATA ATAGCCTCGCCTTTGGTCAGCGTAC TGATAAAAGGTCCGGTGAAATAGTC GCCAGCATCAAAGAAAAAGACATCT TTCTCTTTCGCTTTTGCATCTTTGAC AATTTTCGAGATGGGCGCAAAGCC GCCTACCGGACGTGTCTTGGATACA TAGGGGATAATTTCTGGGGTTACAT G

Sequencing of Promoters.

One hundred and ninety-two clones from a library that underwent two rounds of enrichment in tumor (library-3) were picked at random and sequenced, yielding 100 different sequences. These were mapped to the genome and their potential regulation (tumor-specific activation, or activation in both spleen and tumor) was determined by comparison with the microarray data (see Table 5, presented below). The clones included 26 that were preferentially activated in tumors, and 40 that were activated both in tumor and spleen. 77% of the tumor enriched clones (20 of 26) and 75% of the clones induced in both tumor and spleen (30 of 40) mapped at least partly to intergenic regions. As expected, none of these 100 clones were spleen-specific. The 20 intergenic clones supported by both biological replicates on array experiments are presented in Tables 6A and 6B.

Table 5. Microarry status of active promoter clones in *Salmonella*

Genome Location	Promoter Status		
	Not Detected	Active in Spleen and Tumor	Preferentially Active in Tumor
Intragenic sequences	27	10	6
Intergenic sequences	7	30	20

Table 6A. Cloned candidate intergenic tumor-specific *Salmonella* promoters

Intergenic regions	Genome position of peak signal	Median ratio of experiment versus input				
		Clone ID	Spleen	Tumor (+)	Tumor (+)(-)(+)	Tumor (+)(-)(+)
			Lib-1	Lib-2	Lib-3	Lib-4
STM0468 - STM0469	526177	85	0.9	2.3	5.5	9.5
STM0474 - STM0475	529126	86	1.9	1.7	3.2	2.6
STM0580 - STM0581	638735	87	0.9	3.2	0.3	8.5
STM0844 - STM0845	914762	10	0.8	1.9	5.8	0.4
STM0937 - STM0938	1014704	11	0.7	4.2	6.5	10.3
STM1382 - STM1383	1466034	16	0.7	4.6	7.4	13.9
STM1529 - STM1530	1606103	20	1.9	5.5	2.8	13
STM1807 - STM1808	1909051	26	1.2	1.6	6.5	9.7
STM1914 - STM1915	2011503	28	0.9	3.9	7.2	7.5
STM1996 - STM1997	2079476	30	1.2	2.9	7.4	4
STM2035 - STM2036	2114187	31	1.3	5.9	4.7	8
STM2261 - STM2262	2359663	34	0.6	2.1	3.5	4.8
STM2309 - STM2310	2417301	36	0.6	2.7	6.5	6.3

STM3070 - STM3071	3233025	44	0.8	1.4	2.8	3.1
STM3106 - STM3107	3266543	45	1.1	3.5	4.6	4.6
STM3525 - STM3526	3688646	55	0.8	3.8	1.8	5.6
STM3880 - STM3881	4091492	61	0.9	5.4	0.1	13.8
STM4289 - STM4290	4530650	71	0.9	2	8.3	10
STM4418 - STM4419	4661108	77	0.8	3.4	8.3	6
STM4430 - STM4431	4674477	78	1.3	6.1	5.6	8

Table 6B. Cloned candidate intergenic tumor-specific *Salmonella* promoters (cont'd)

Intergenic regions	Clone ID	Cloned Promoter	5' gene	5' gene orient.	3' gene	3' gene orient.	Anerobic induction?	Stable / Unstable GFP
STM0468 - STM0469	85	+	ylaB	-	rpmE2	+		Unstable
STM0474 - STM0475	**86**	-	ybaJ	-	acrB	-	.	Stable
STM0580 - STM0581	**87**	-	STM0580	-	STM0581	+		Stable
STM0844 - STM0845	**10**	-	pflE	-	moeB	-	Yes	Unstable
STM0937 - STM0938	**11**	-	hcp	-	ybjE	-	Yes	Unstable
STM1382 - STM1383	**16**	-	orf408	-	ttrA	-		Stable
STM1529 - STM1530	**20**	-	STM1529	+	STM1530	+		Stable
STM1807 - STM1808	**26**	+	dsbB	+	STM1808	+		Stable
STM1914 - STM1915	**28**	-	flhB	-	cheZ	-		Unstable
STM1996 - STM1997	**30**	-	cspB	-	umuC	-		Stable
STM2035 - STM2036	**31**	-	cbiA	-	pocR	-		Stable
STM2261 - STM2262	**34**	-	napF	-	eco	+	Yes	Stable
STM2309 -	**36**	-	menD	-	menF	-		Stable

STM2310								
STM3070 - STM3071	**44**	-	epd	-	STM3071	+		Unstable
STM3106 - STM3107	**45**	-	ansB	-	yggN	-	Yes	Stable
STM3525 - STM3526	**55**	+	glpE	+	glpD	+		Stable
STM3880 - STM3881	**61**	+	kup	+	rbsD	+		Stable
STM4289 - STM4290	**71**	-	phnA	-	proP	+		Unstable
STM4418 - STM4419	**77**	+	STM4418	-	STM4419	+		Stable
STM4430 - STM4431	**78**	+	STM4430	-	STM4431	+		Stable

Some possible tumor promoters mapped inside annotated genes; 23% of the sequenced clones (6 of 26) and 18% of candidates identified by microarray (19 of 105; see Table 7, presented below). Some "promoters" may be artifacts that could arise from a variety of effects such as the inherent high copy number of the plasmid clone, or mutations that cause the copy number to increase or a new promoter to be generated. However, based on data from *Escherichia coli* , a close relative of *Salmonella*, intragenic regions might indeed contain promoters, based on evidence from transcription start sites, binding sites for RNA polymerase (Reppas et al, "The transition between transcriptional initiation and elongation in *E. coli* is highly variable and often rate limiting", Mol. Cell 24:747-757, 2006, Grainger et al, "Studies of the distribution of *Escherichia coli* cAMP-receptor protein and RNA polymerase along the *E. coli* chromosome", Proc. Natl. Acad. Sci. USA 102:17693-17698, 2005), and sigma factors (Wade et al, "Extensive functional overlap between sigma factors in *Escherichia coli*", Nat. Struct. Mol. Biol. 13:806-814, 2006) as well as motif finders (Tutukina et al, "Intragenic promoter-like sites in the genome of Eschericia coli discovery and functional implication", J. Bioinform. Comput. Biol. 5:549-560, 2007). Further work may provide confirmatory evidence of promoter activity in some cases.

Some weaker promoters may generate detectable GFP in the stable, but not the destabilized, GFP plasmid library. Fifty clones sequenced after FACS selection could be assigned to either the stabilized or destabilized library. Forty of these were of the stable GFP variety versus an expected

25 of each type if there had been no bias. Therefore, the destabilized library is, as expected, underrepresented following FACS.

Table 7. Intragenic regions that induce higher GFP expression in tumor than in spleen

Clone ID	Spleen	Tumor (+)	Tumor (+)(-)(+)	Tumor (+)(-)(+)	Genome position of peak signal	Gene	Gene symbol			
	lib1	lib2	lib3	lib4						
	Median of experiment versus input library							intragenic seq. orient.	Gene orient	
Seq'd										
1	0.64	3.16	4.47	3.01	40,802	STM0035	STM0035	-	+	CCCGCGCTATGGCGTGGT GCATCCTACGGGGTGGAT TCGTAATGGCCAACATATT GGCCGCGCAGATAAGATG AGCGGCGAGTTTGTGAGC TCTGAAGTGGTGAACTGG CTGGATAATAAGAAAGACG ATAATCCGTTCTTCTTATAT GTCGCCTTTACCGAAGTCC ATAGCCCGCTGGCGTCGC CGAAAAAATACCTTGATAT GTATTCGCAGTACATGACC GACTACCAGAAGCAGCAT CCGGATCTGTTCTACGGC GACTGGGCAGACAAACCG TGGCGCGGCACCGGCGAA TATTAC

84	0.61	1.48	3.99	2.76	558,116	STM0498	ybaR	-	-	CAATAGCCGGTTGGCATTG CTGACGACGGTAATGGAA GACAGCGCCATTGCCGCG CCTGCTACTACCGGGTTTA ACAAGGTACCGGTAAACG GCCACAGAATACCGGCGG CCACCGGGATACCAATGC TGTTGTAGATAAATGCGCC AAGCAGGTTTTGTTTCATA TTGCGCAACGTCGCGCGC GAAATGGCCAGCGCATCC GCCACGCCCATCAGACTAT GGCGCATCAGCGTAATCG CCGCGGTTTCAATCGCCA CATCGCTGCCGCCGCCCA TCGCGATACCGACGTCCG CCTGCGCC
7	0.68	6.89	4.77	10.76	743,461	STM0683	nagA	-	-	TAGTCGACATGCAGACCAT CGGCGATAACGCCGCAAT AAATATCCGCTTCGTCCAG AACAGCGCCAGCAAGGCC CGGCTCACGCCCTGTAAT GTACGGCATCGCGTTAAAC AGGTGAGTCGCAAAGGTA ATCCCGGCGCGGAAGCCC GCTTTCGCCTCTTTTAACG TCGCGTTGGAGTGACCTG CGGAAACCACAATGCCCG CATTCGCCAGTTTAGCGAT TACGTCAGCAGGCACCATT TCCGGCGCGAGTGTGACT TTGGTGATGACGTCGGCAT TATCGCATAAGAAATCGAC CAGCG

15	0.73	6.11	0.24	14.71	1,418,744	STM1338	pheT	+	+	ATGAATCCGGCTCTGCATC CGGGACAGTCTGCGGCGA TTTATCTGAAAGATGAACG TATTGGTTTTATTGGGGTT GTTCACCCTGAACTGGAAC GTAAACTGGATCTGAATGG TCGTACGCTGGTGTTTGAA CTGGAATGGAATAAGCTCG CAGACCGTATCGTGCCGC AGGCGCGGGAGATTTCAC GCTTCCCGGCCAACCGTC GCGATATTGCGGTTGTTGT TGCAGAAAACGTTCCCGCA GCGGATATTTTATCCGAAT GTAAGAAAGTTGGCGTAAA TCAGGTAGTTGGCGTAAAC T
17	0.83	3.46	3.23	5.23	1,504,175	STM1426	ribE	+	+	CGTGCATCTCATTCCGGAA ACGTTGGAACGTACTACGC TTGGCAGAAAAAAACTGGG TGAGCGTGTGAATATCGAG ATCGATCCGCAAACGCAG GCGGTTGTCGATACCGTA GAACGCGTACTGGCTGCG CGAGAAAATGCGGTCAGA AATCAGGCCGACATTGGCT AACGGAAAATAAGATTCCC CCGCATGAAATGCGGGGG AGATGATTAGCGAGGAAC GCGCAGTCCGTTTTCAACG CCGCGCGTAAATACCACCT GCCAAAGCTGGATATCAC GCGCGCGAAACGCACCCG CGCAG

56	0.70	6.90	4.49	23.58	3,523,313	STM3355	STM3355	+	-	TTTCAACAGAGGTCGCTAC GCCCACGCCAACCAGCAG CGGCGGACAAGCGTTGAG GCCGTAGCTGGTCATCAC ATCCAGTACAAAGCGGGT CACACCTTCATAGCCTGCA CCCGGCATCAGCACCATC GCTTTCCCCGGCAGAGAA CAACCACCGCCCGCCATA TAGGTATAAATGCTGCACT GATCGGAATTGGGAACGA TTTCCCAGAAGACCGTCG GCGTACCTTTACCCACGTT TTTACCGGTGTTGTATTCA TCAAAAGTTTCTACGCTGT TGTGGCGCAGCGGAGAAT CTACAGT
array data only										
	0.91	7.43	3.70	5.41	18,084	STM0018	STM0018			ACCCTGCAACAAGCGATG ATAAAATAAGCTATACTTC CGCTCGTGTTGCGAAACC GGTATACAATAAATATAAA AATTCCACGACTAAACCGA AGGTATTCGGTTATTACAC CGACTGGTCACAGTATGAC AGCCGTCTGCAAGGCAAT ATGTCCCAACCGGGCCGT GGTTATGATTTAACCAAAG TTTCACCGACGGCTTATGA CAAACTGATTTTTGGCTTT GTTGGCATCACCGGTTTCA GAAAAATTGATACAGAAGA CCGCGATGTCGTAGCAGA AGCGGCAGCGCTGTGCGG CAA

	0.92	2.12	4.85	6.29	1,071,228	STM0984	msbA			AAGAGGTACTGATTTTTGG CGGTCAGGAAGTCGAAAC TAAACGCTTTGATAAAGTC AGCAATAAGATGCGACTGC AAGGCATGAAAATGGTCTC TGCCTCGTCAATTTCCGAT CCTATCATTCAGCTCATTG CCTCGCTGGCGCTGGCGT TTGTCCTCTATGCTGCGAG CTTCCCAAGCGTAATGGAT AGCCTGACGGCAGGGACC ATCACCGTGGTGTTCTCCT CCATGATCGCGCTGATGC GTCCATTAAAATCGCTGAC AAACGTTAACGCGCAGTTC CAGCGTGGGATGGCGGCT TG
	0.46	3.08	2.56	4.03	1,342,729	STM1258	STM1258			GCGCGAGACGCTGGTCGC CGTTATTACAGAATGTCTC TTTTGATATCGCGCCCGGC GAAATGGTGGCATTGGTTG GCGGCAGCGGGGAGGGC AAAAGTCTGCTGCTGCAAT GCCTGCTCGATCTGCTGC CGGAAAATTTACGCTTTCG GGGGGAGATTACGCTTGA TGGCAACCGGCTGGACAG ACATACCATCAGGCAGCTT AGGGGCAATACGTTTAGCT ACGTGCCGCAGGGGGTAC AGGCGCTTAATCCCATGCT GAATATCAGAAAACATTTG AACAGAGCATGTCATCTGA CCGG

	0.91	2.09	3.01	4.08	2,358,604	STM2259	napA			ATTGACCCGATCCAAACAT GCCGATCGCTTCTGGTCCT TTCTCTTTCAGGGAGGTTT TAAACTTCTCTTCCATCAC ATCGAAGGCCTGTTCCCA GCTCACCGGCGTAAACTC GCCGTCTTTGTGATAGCTG CCGTCTTTCATGCGCAGCA TCGGCTGCGTCAGACGAT CTTTACCGTACATGATTTT GGGCAGGAAGTAGCCTTT AATGCAGTTCAGACCACG GTTGACCGGCGCGTCGGG GTCGCCCTGGCAGGCGAC CACACGGCCCTGCTGCGT TCCCACCAACACACCGCAA CCCGT
	1.40	2.88	3.62	9.57	3,002,027	STM2857	hypD			CACATTACGCTGATCCCGA CGCTGCGTAGCCTACTGG AGCAGCCGGACAACGGCA TTGACGCCTTTCTTGCGCC AGGCCACGTCAGCATGGT CATCGGCACCGAGGCGTA CCAGTTTATCGCCGCCGAT TTTCATCGCCCGCTGGTG GTGGCTGGATTCGAACCG CTTGATCTACTGCAAGGCG TGGTCATGCTGGTTGAGCA GAAAATAGCGGCCCTAAG CCAGGTTGAAAATCAATAC CGTCGCGTGGTGCCGGAT GCCGGAAACATGCTGGCG CAGCAGGCCATTGCCGAT GTGTTCT

	0.74	2.66	7.94	22.93	3,026,126	STM2882	sipA			AGCAGCAGGGGTATCAAC GTTTGCATTTCAAGGTGCC GGGCTTCCCGTCCTACGC TGGTACCCTGCTCTTGCGT TAATTTTTGGTGGCACATA TCAAGCGCCTCAACAGCCT TCGCCGCCGCTTTGTCAAC AAGGTGCGTAAGATTGCTG CGGGTTAACGGATCTAAC GTACAGCCAAAGTTATGTT CAATGCAGCTGGCAATATA GGGCATCACCTCCTGCATA ACAAGATTCGTCGATAATT TACTTAATTCACCGCCAGT GTTATTTTTGATAATATCTA ACAGCTGCTTTTCCAGGT
	0.74	3.02	5.85	17.96	3,087,704	STM2945	sopD			TAGAATCTATGAGTAGAGA GGAGAGACAATTATTTTTA CAAATATGTGAGGTGATTG GTTCGAAGATGACCTGGC ACCCGGAATTACTTCAGGA GTCGATTTCAACTCTACGA AAAGAAGTGACGGGAAAT GCACAAATCAAAACGGCG GTTTATGAGATGATGCGTC CCGCAGAGGCTCCAGACC ACCCGCTTGTCGAATGGC AGGACTCACTTACTGCAGA TGAAAAATCAATGCTGGCC TGTATTAATGCCGGTAACT TTGAGCCTACGACTCAGTT TTGCAAAATAGGTTATCAG GA

	0.81	3.08	3.19	7.02	3,472,959	STM3304	rplU			GTGAACCACTGACGATGG CCCTGCTGCTTACGGTAGT GTTTACGGCGACGAAACTT AACGATTTTAACTTTCTCG CCACGACCGTGGGCAACA ACTTCAGCTTTGATTACGC CGCCATCAACGAAAGGAA CGCCGATTTTGACTTCTTC ACCGTTTGCGATCATCAGA ACTTCAGCGAACTCGATAG TTTCGCCAGTTGCGATGTC CAGCTTTTCCAGGCGAAC GGTCTGACCTTCGCTTACT CGGTGTTGTTTACCACCAC TTTGGAAAACCGCGTACAT AAAAAACTCCGCTTCCGCG C
	0.73	2.63	2.53	5.18	3,660,088	STM3502	ompR			CGCCGGGCAGTTCGTTTG CCTGACGACGTAACACGG CGCGAATACGCGCCAACA GCTCGCGCGGGTTAAACG GTTTAGGAATGTAGTCATC GGCGCCGATTTCCAGCCC GACGATACGGTCAACCTCT TCACCCTTCGCCGTGACCA TAATGATCGGCATTGGATT ACTTTGACTACGCAGGCGA CGACAAATCGACAGACCAT CTTCACCTGGCAGCATTAA ATCCAGTACCATGAGATGG AAAGATTCACGGGTCAGCA GACGATCCATCTGCTCAGC GTTAGCGACGCTTCGAAC CTG

	0.89	3.00	3.86	3.92	3,957,871	STM3758	fidL			GCTTAATGCGTACAGAAAA ATATCGGGCGTTTCCCGAT GGTGAACATAAAGCCACG ATGGCCCTGAGTCAGGAT GGTGTAACTGATACTTTTC CCTGGATAGACATAAAAAT CGGGTAAAACCGTCTCGAT AACCGCATCGGACAGTGTT TCGTCACGCGTGACTTTGT TGATATCCGTCGATATAAA ATGGGTGCTGTCTTTATTT TCACTCCATACATAGGAAA CATCACGGCGGATCACGC CGCTCATTTTATTATCGAC GTAATATGTTCCGCTGATG GAAACCACCCCAGTGCGT T
	0.73	7.03	2.38	11.84	4,601,412	STM4358	amiB			CCGAACTGTTAGGCGGCG CTGGCGATGTGCTGGCGA ACAGTCAGTCAGACCCTTA CCTGAGCCAGGCGGTACT GGATTTGCAATTCGGTCAT TCGCAGCGGGTAGGGTAT GATGTGGCGACGAACGTA CTAAGCCAACTCGACGGC GTGGGGTCGCTGCATAAA CGCCGCCCGGAACACGCT AGCCTGGGCGTGTTGCGT TCGCCGGATATCCCGTCC ATTTTGGTGGAGACGGGC TTTATCAGTAATCACGGCG AAGAGCGATTGCTGGCGA GCGACCGCTATCAGCAGC AGATTGCTGA

	0.49	5.44	8.71	19.81	4,735,184	STM4489	STM4489			TTTCCTGAATCAGACGTTT GAAAATACCGATAAACACA TCACGATAGTTTCTCCATG GCTAACCTGGCAAAAACTG GAGCAAACCGGTTTTCTTG ATTCCATGATTACGGCGTG TTCACGTGGTATTAACGTC ACGGTAGTCACTGACAGAA GCTACAACACTGAACATAA TGATTTTGAGAAGCGAAAA GAGAAGCAGCAGAACCTT AAAGCGGCGCTGGAGAAA CTGAACGCCCTTGGTATTG CGACAAAACTGGTCAATCG TGTTCATAGCAAAATTGTT ATTGGTGATGATGGTTTG
	0.64	11.20	6.44	19.39	4,748,275	STM4496	STM4496			TTTGCGCGCCAGACGGGC AACCAGCAGCTTCACTTCT TCTTCCGGCCATCCATAAG GACGGCGGGCAAAGTGGT TCAGAATATCGCGTAAATA AACCGGCTTATTGAACTCG ATATTCATGCTGACCCAGG TTTCTACTTCGCGCATCGC GTCGGGGTTGGATTCCTC CAGTTCGCCCAGATCCAG CTCCGCATCATTCTCCACC GTGAGTAGTGCATGGATTT CACGTGCGATATCACCGTT GAACGGGCGCAGCATTTT CAGCTTGGCAAACGTGTTT TCAATCACATAGCGGCAAG CT

Confirmation of tumor specificity of individual clones *in vivo*.

Five cloned promoters potentially activated in bacteria growing in tumor but not in the spleen were selected to be individually confirmed *in vivo*. A group of tumor-bearing mice and normal mice were injected i.v. with bacteria containing the cloned promoters. Tumors and spleens were imaged after 2 days, at low and high resolution using the Olympus OV 100 small animal imaging system. Three of the five tumor-specific candidates (clones 10, 28, and 45) were induced much more in tumor than in spleen. Clone 44 produced low signals and clone 84 was highly expressed in tumor but was detectable in the spleen.

Among the most likely promoters to be uncovered in this study are those induced by hypoxia, which is thought to be an important contributor to *Salmonella* targeting of tumors (Mengesha et al, "Development of a flexible and potent hypoxia-inducible promoter for tumor-targeted gene expression in attenuated *Salmonella*", Cancer Biol. Ther. 5:1120-1128, 2006). *Salmonella* promoters induced by hypoxia include those controlled directly or indirectly by the two global regulators of anaerobic metabolism, Fnr and ArcA (Iuchi and Weiner, Cellular and molecular physiology of *Escherichia coli* in the adaptation to aerobic environments", J. Biochem. 120:1055-1063, 1996).

Clone 45 contains the promoter region of *ansB* , which encodes part of asparaginase. In *E. coli*, *ansB* is positively coregulated by Fnr and by CRP (cyclic AMP receptor protein), a carbon source utilization regulator (24). In *S. enterica*, the anaerobic regulation of *ansB* may require only CRP (Jennings et al, "Regulation of the *ansB* gene of *Salmonella enterica*", Mol. Miicrobiol. 9:165-172, 1993, Scott et al, "Transcriptional co-activation at the *ansB* promoters: involvement of the activating regions of CRP and FNR when bound in tandem", Mol. Microbiol. 18:521-531, 1995).

Clone 10 is the promoter region of a putative pyruvate-formate-lyase activating enzyme (*pflE*). This clone was only observed in library-3, but enrichment was considerable in that library (see Tables 2A and 2B). This clone was pursued further because the operon is co-regulated in *E. coli* by both ArcA and Fnr (Sawers and Suppmann, "Anaerobic induction of pyruvate formate-lyase gene expression is mediated by the ArcA and FNR proteins", J. Bacteriol. 174:3474-3478, 1992, Knappe and Sawers, "A radical-chemical route to acetyl-CoA: the anaerobically induced pyruvate formate-lyase system of *Escherichia coli*", FEMS Microbiol. Rev. 6:383-398, 1990).

Finally, clone 28 contains the promoter region of *flhB*, a gene that is required for the formation of the flagellar apparatus (Williams et al, "Mutations in *fliK* and *flhB* affecting flagellar hook and filament assembly in *Salmonella typhimurium*" J. Bacteriol. 178:2960-2970, 1996) and is not known to be regulated in anaerobic metabolism.

Further screening was performed on these three clones. Bacteria containing these clones were i.v. injected at 5 x 10^6, 5 x 10^7, and 5 x 10^7 cfu into tumor and non-tumor-bearing nude mice. One or 2 days post-injection, spleens and tumors were imaged using the OV100 imaging system, homogenized, and the bacterial titer was quantified on LB+Amp. Spleens from normal mice were compared with tumors that had a similar number of colony-forming units, so that any difference in fluorescence would be attributable to increased GFP expression rather than bacterial numbers. FIG. 2 confirms that tumors are much more fluorescent than spleens infected with the same number of bacteria for each of the three clones. A positive control that constitutively expresses TurboGFP resulted in strong fluorescence in spleen even with doses as low as 2 x 10^5 cfu.

The *Salmonella* endogenous promoter for *pepT* is regulated by CRP and Fnr (Mengesha et al, 2006). In previous studies, the TATA and the Fnr binding sites of this promoter were modified to engineer a hypoxia-inducible promoter that drives reporter gene expression under both acute and chronic hypoxia *in vitro* (Mengesha et al, 2006). Induction of the engineered hypoxia-inducible promoter *in vivo* became detectable in mice 12 hours after death, when the mouse was globally hypoxic (Mengesha et al, 2006). In our experiments, the wild-type *pepT* intergenic region did not pass the threshold to be included in the tumor-specific promoter group. Perhaps the appropriate clone is not represented in the library, or induction (i.e., level of hypoxia in the PC3 tumors) was not enough for this particular promoter.

In summary, *Salmonella* thrives in the hypoxic conditions found in solid tumors (Mengesha et al, 2006). There are four promoters known to be regulated by hypoxia among the 20 sequenced intergenic clones (see Tables 2A and 2B), of which two (clones 10 and 45) were tested and shown to be induced in tumors (see FIG. 2). Many candidate promoters that seem to be preferentially activated within tumors may be unrelated to hypoxia, including clone 28 (FIG. 2). Any promoters that are later proven to respond in their natural context in the genome may illuminate conditions within tumors, other than hypoxia, that are sensed by *Salmonella*.

Attenuated *Salmonella* strains with tumor targeting ability can be used to deliver therapeutics under the control of promoters preferentially induced in tumors (Pawelek et al. "Tumor-targeted *Salmonella* as a novel anticancer vector", Cancer Res 1997; 57:4537-44; Zhao et al. "Targeted therapy with a *Salmonella* typhimurium leucine-arginine auxotroph cures orthotopic human breast tumors in nude mice", Cancer Res 2006; 66:7647-52; Zhao et al. "Tumor-targeting bacterial therapy with amino acid auxotrophs of GFP-expressing *Salmonella* typhimurium", Proc Natl Acad Sci U S A 2005; 102:755-60; Zhao et al. "Monotherapy with a tumor-targeting mutant of *Salmonella* typhimurium cures orthotopic metastatic mouse models of human prostate cancer", Proc Natl Acad Sci U S A 2007; Nishikawa et al. "*In vivo* antigen delivery by a *Salmonella* typhimurium type III secretion system for therapeutic cancer vaccines", J Clin Invest 2006; 116:1946-54; Panthel et al. "Prophylactic anti-tumor immunity against a murine fibrosarcoma triggered by the *Salmonella* type III secretion system", Microbes Infect 2006; 8:2539-46; Thamm et al. "Systemic administration of an attenuated, tumor-targeting *Salmonella* typhimurium to dogs with spontaneous neoplasia: phase I evaluation", Clin Cancer Res 2005; 11:4827-34; Forbes et al. "Sparse initial entrapment of systemically injected *Salmonella* typhimurium leads to heterogeneous accumulation within tumors", Cancer Res 2003; 63:5188-93; Toso et al. "Phase I study of the intravenous administration of attenuated *Salmonella* typhimurium to patients with metastatic melanoma", J Clin Oncol 2002; 20:142-52; Avogadri, et al. "Cancer immunotherapy based on killing of *Salmonella*-infected tumor cells", Cancer Res 2005; 65:3920-7). Such promoters are technically useful whether or not they are regulated in the same way in their natural context in the genome. These promoters would be tools to reduce the expression of the therapeutic in bacteria outside the tumor and thus reduce side-effects, and thereby produce a highly selective and effective therapy of metastatic cancer. Further sophistications are also possible. For example, combinations of two or more promoters that are preferentially induced in tumors by differing regulatory mechanisms would allow delivery of two or more separate protein components of a therapeutic system under different regulatory pathways. In addition, new promoter systems induced by external agents such as arabinose (Loessner et al. "Remote control of tumor-targeted *Salmonella enterica* serovar Typhimurium by the use of L-arabinose as inducer of bacterial gene expression *in vivo*", Cell Microbiol. 9:1529-37, 2007) or salicylic acid (Royo et al. "*In vivo* gene regulation in *Salmonella* spp. by a salicylate-dependent control circuit", Nat. Methods 4:937-42, 2007) allow promoters in *Salmonella* to be induced throughout the body at a time of choice. Such inducible regulation could be combined with tumor-specific *Salmonella* promoters to express useful products in the tumor only when the exogenous activator is added; therapy delivery would be exquisitely controlled both in time and space.

* * *

The entirety of each patent, patent application, publication and document referenced herein hereby is incorporated by reference. Citation of the above patents, patent applications, publications and documents is not an admission that any of the foregoing is pertinent prior art, nor does it constitute any admission as to the contents or date of these publications or documents.

Modifications may be made to the foregoing without departing from the basic aspects of the invention. Although the invention has been described in substantial detail with reference to one or more specific embodiments, those of ordinary skill in the art will recognize that changes may be made to the embodiments specifically disclosed in this application, yet these modifications and improvements are within the scope and spirit of the invention.

The invention illustratively described herein suitably may be practiced in the absence of any element(s) not specifically disclosed herein. Thus, for example, in each instance herein any of the terms "comprising," "consisting essentially of," and "consisting of" may be replaced with either of the other two terms. The terms and expressions which have been employed are used as terms of description and not of limitation, and use of such terms and expressions do not exclude any equivalents of the features shown and described or portions thereof, and various modifications are possible within the scope of the invention claimed. The term "a" or "an" can refer to one of or a plurality of the elements it modifies (e.g., "a reagent" can mean one or more reagents) unless it is contextually clear either one of the elements or more than one of the elements is described. The term "about" as used herein refers to a value within 10% of the underlying parameter (i.e., plus or minus 10%), and use of the term "about" at the beginning of a string of values modifies each of the values (i.e., "about 1, 2 and 3" refers to about 1, about 2 and about 3). For example, a weight of "about 100 grams" can include weights between 90 grams and 110 grams. Further, when a listing of values is described herein (e.g., about 50%, 60%, 70%, 80%, 85% or 86%) the listing includes all intermediate and fractional values thereof (e.g., 54%, 85.4%). Thus, it should be understood that although the present invention has been specifically disclosed by representative embodiments and optional features, modification and variation of the concepts herein disclosed may be resorted to by those skilled in the art, and such modifications and variations are considered within the scope of this invention.

Certain embodiments of the invention are set forth in the claims that follow.

What is claimed is:

1. An isolated nucleic acid molecule which comprises a recombinant expression system, which expression system comprises a nucleotide sequence encoding a toxic or therapeutic RNA or protein, or an RNA or protein that participates in generating a toxin or therapeutic agent operably linked to a heterologous promoter which promoter is preferentially activated in solid tumors.

2. The isolated nucleic acid molecule of claim 1 wherein the promoter is an Enterobacteriaceae promoter.

3. The isolated nucleic acid molecule of claim 2 wherein the promoter is a Salmonella promoter.

4. The isolated nucleic acid molecule of claim 3, wherein the promoter comprises (i) a nucleotide sequence of Table 7A and Table 7B, or (ii) a functional promoter subsequence of (i).

5. The isolated nucleic acid molecule of claim 4, wherein the functional promoter subsequence is about 20 to about 150 nucleotides in length.

6. Recombinant host cells that contain the nucleic acid molecule of any of claims 1-5.

7. The cells of claim 6 that are avirulent Salmonella.

8. A pharmaceutical composition which comprises the nucleic acid molecule of claims 1-5 or the cells of claims 6-7.

9. A method to treat solid tumors which method comprises administering to a subject harboring said tumors the nucleic acid molecule of claims 1-5 or the cells of claims 6-7 or the composition of claim 8.

10. A method for identifying a promoter preferentially activated in tumor tissue which method comprises:

(a) providing a library of expression systems each comprising a nucleotide sequence encoding a detectable protein operably linked to a different candidate promoter;

(b) providing said library to solid tumor tissue and to normal tissue;

(c) identifying cells from each tissue that show high levels of expression of the detectable protein; and

(d) obtaining said expression systems from the cells that produce greater levels of detectable protein in tumor tissue as compared to normal tissue, and identifying the promoters of said expression system.

11. The method of claim 10 wherein said library is provided in recombinant host cells.

12. The method of claim 10 or claim 11 wherein the promoters are Salmonella promoters and the recombinant host cells are Salmonella.

13. The method of any one of claims 10-12, wherein the candidate promoters are from bacteria, or are 80% or more identical to promoters from bacteria.

14. The method of claim 13, wherein the bacteria are Enterobacteriaceae.

15. The method of claim 14, wherein the Enterobacteriaceae are Salmonella.

16. The method of any one of claims 10-15, which comprises scoring promoters identified in (d).

17. An expression system which comprises a nucleotide sequence encoding a toxic or therapeutic protein or a protein that participates in generating a desired toxin or therapeutic agent operably linked to a promoter identified by the method of any of claims 10-16.

18. Recombinant host cells that comprise the expression system of claim 17.

19. A method to treat solid tumors which method comprises administering an expression system of claim 17 or the cells of claim 18 to a subject harboring a solid tumor.

20. The method of claim 19, wherein the protein encoded by the nucleotide sequence comprises enzymic activity.

21. The method of claim 20, which comprises administering a prodrug to the subject that does not inhibit tumors, wherein the protein encoded by the nucleotide sequence coverts the prodrug to a drug that inhibits tumors.

22. An expression system which comprises a first promoter nucleotide sequence operably linked to a first coding sequence and second promoter nucleotide sequence operably linked to a second coding sequence, wherein:

the first coding sequence and the second coding sequence encode polypeptides that individually do not inhibit tumor growth;

polypeptides encoded by the first coding sequence and the second coding sequence, in combination, inhibit tumor growth; and

the first promoter nucleotide sequence and the second promoter nucleotide sequence are preferentially activated in solid tumors.

23. The expression system of claim 22, wherein the first promoter nucleotide sequence and the second promoter nucleotide sequence are in the same nucleic acid molecule.

24. The expression system of claim 22, wherein the first promoter nucleotide sequence and the second promoter nucleotide sequence are in different nucleic acid molecules.

25. The expression system of any one of claims 22-24, wherein the first promoter nucleotide sequence and the second promoter nucleotide sequence are bacterial nucleotide sequences.

26. The expression system of claim 25, wherein the bacterial sequences are Enterobacteriaceae sequences.

27. The expression system of claim 26, wherein the Enterobacteriaceae sequences are Salmonella sequences.

28. The expression system of any one of claims 22-27, wherein:

the first coding sequence encodes an enzyme,

the second coding sequence encodes a prodrug, and

the enzyme processes the prodrug into a drug that inhibits tumor growth.

29. The expression system of any one of claims 22-27, wherein:

the first coding sequence encodes a first polypeptide,

the second coding sequence encodes a second polypeptide, and

the first polypeptide and the second polypeptide form a complex that inhibits tumor growth.

30. The expression system of any one of claims 22-30, wherein the first promoter nucleotide sequence, the second promoter nucleotide sequence, or the first promoter nucleotide sequence and the second promoter nucleotide sequence comprise (i) a nucleotide sequence of Table 7A and Table 7B, (ii) a functional promoter nucleotide sequence 80% or more identical to a nucleotide sequence of Table 7A and Table 7B, or (iii) or a functional promoter subsequence of (i) or (ii).

31. The expression system of claim 30, wherein the functional promoter subsequence is about 20 to about 150 nucleotides in length.

32. Recombinant host cells that contain the expression system of any one of claims 22-31.

33. The cells of claim 32 that are avirulent Salmonella.

34. An expression system which comprises three or more heterologous promoter nucleotide sequences operably linked to three or more coding sequences, wherein the promoter nucleotide sequences are preferentially activated in solid tumors.

35. The expression system of claim 34, wherein the coding sequences encode polypeptides that individually do not inhibit tumor growth, and the polypeptides encoded by the coding sequences, in combination, inhibit tumor growth.

36. The expression system of claim 34 or 35, wherein the promoter nucleotide sequences are in the same nucleic acid molecule.

37. The expression system of claim 34 or 35, wherein the promoter nucleotide sequences are in different nucleic acid molecules.

38. The expression system of any one of claims 34-37, wherein the promoter nucleotide sequence are bacterial nucleotide sequences.

39. The expression system of claim 38, wherein the bacterial sequences are Enterobacteriaceae sequences.

40. The expression system of claim 39, wherein the Enterobacteriaceae sequences are Salmonella sequences.

41. The expression system of any one of claims 34-40, wherein the first promoter nucleotide sequence, the second promoter nucleotide sequence, or the first promoter nucleotide sequence and the second promoter nucleotide sequence comprise (i) a nucleotide sequence of Table 7A and Table 7B, (ii) a functional promoter nucleotide sequence 80% or more identical to a nucleotide sequence of Table 7A and Table 7B, or (iii) or a functional promoter subsequence of (i) or (ii).

42. The expression system of claim 41, wherein the functional promoter subsequence is about 20 to about 150 nucleotides in length.

43. Recombinant host cells that contain the expression system of any one of claims 34-42.

44. The cells of claim 43 that are avirulent Salmonella.

PATENT
VIV-1001-PC

METHODS TO TREAT SOLID TUMORS

Abstract

A high throughput method for identifying promoters differentially activated in solid tumors as compared to normal tissues is described. The promoters so identified may be used to drive production of RNA's or proteins useful in treating solid tumors including toxic RNA's or proteins and other therapeutic RNA's or proteins.

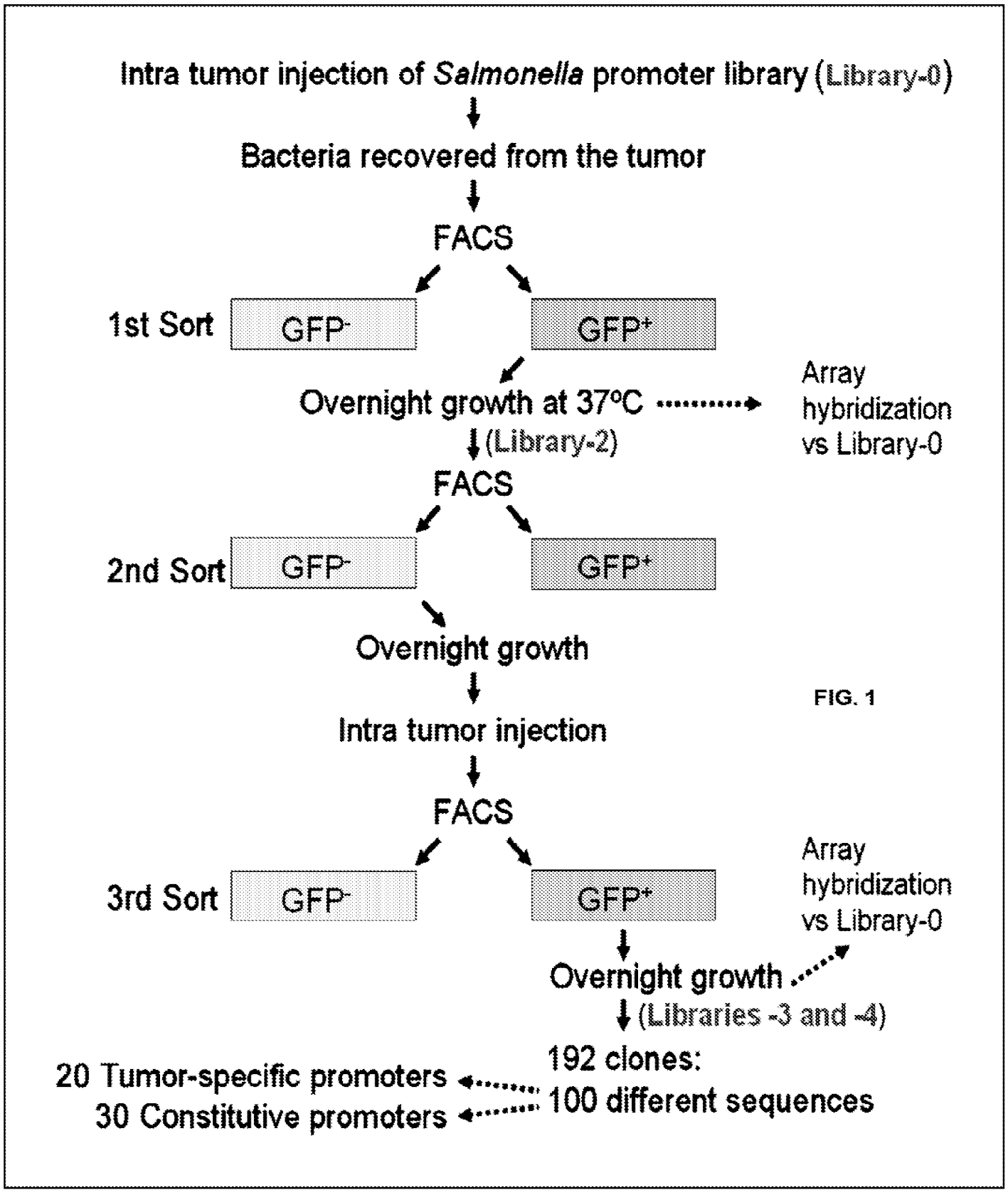

Intra tumor injection of *Salmonella* promoter library (Library-0)
Bacteria recovered from the tumor
FACS
1st Sort
GFP-
GFP+
Overnight growth at 37°C
(Library-2)
Array hybridization vs Library-0
FACS
2nd Sort
GFP-
GFP+
Overnight growth
Intra tumor injection
FACS
3rd Sort
GFP-
GFP+
Array hybridization vs Library-0
Overnight growth
(Libraries -3 and -4)
192 clones:
100 different sequences
20 Tumor-specific promoters
30 Constitutive promoters

FIG. 1

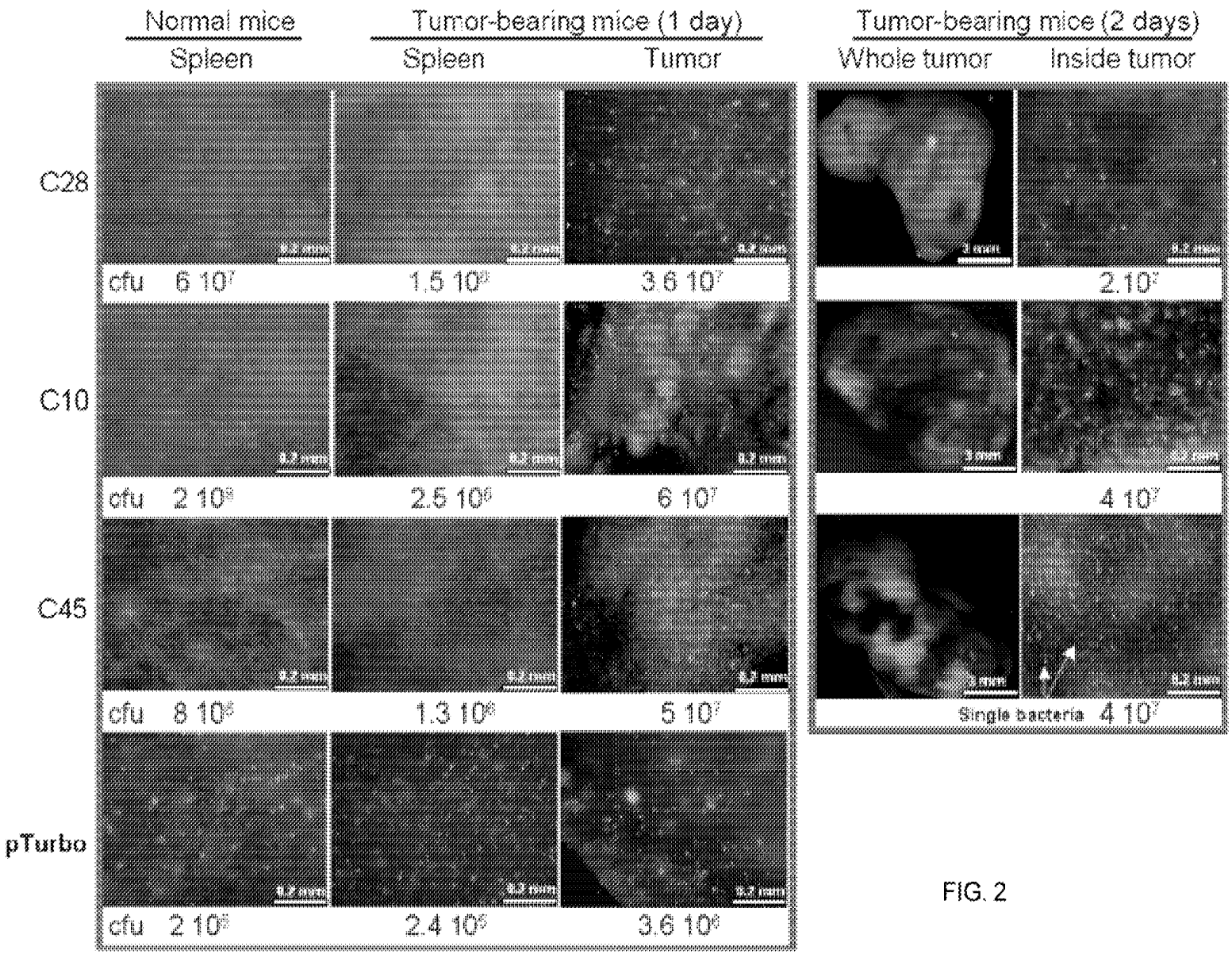

Normal mice
Spleen
Tumor-bearing mice (1 day)
Spleen
Tumor
Tumor-bearing mice (2 days)
Whole tumor
Inside tumor
C28
cfu 6 10⁷
1.5 10⁸
3.6 10⁷
2.10⁷
C10
cfu 2 10⁸
2.5 10⁸
6 10⁷
4 10⁷
C45
cfu 8 10⁸
1.3 10⁸
5 10⁷
Single bacteria 4 10⁷
pTurbo
cfu 2 10⁸
2.4 10⁸
3.6 10⁸

FIG. 2